What information is represented in the human hippocampus?

Martin James Chadwick

Submitted for PhD in Cognitive Neuroscience

May 2012

Supervisor: Eleanor A. Maguire

I, Martin Chadwick, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Date:

Abstract

The hippocampus plays a critical role in supporting memories of our personal past experiences (episodic memories). However, it is not known how individual episodic memories are represented by neuronal populations within the hippocampus. The aim of my thesis was to explore the nature of the information represented in the human hippocampus, with a particular focus on episodic memories.

I conducted five experiments using high-resolution and standard functional MRI (fMRI). In four of these projects I used and further developed a method known as multi-voxel pattern analysis (MVPA). This enabled me to interrogate the fMRI data to look for functionally-relevant patterns of information encoded across multiple voxels. My findings revealed that episodic memories were represented in the hippocampus more so than in neighbouring brain regions, that this was true even of memories that were highly overlapping in terms of content and context, and for recently-formed and very remote memories. Furthermore, I found that the episodic information within individual hippocampal subfields was consistent with computational models of episodic memory.

One important contribution to the representation of an episodic memory is scene construction - the mental construction of a complex spatially coherent scene into which event details are bound. In order to explore the role of the hippocampus in scene construction, I used fMRI to study boundary extension – a scene-related phenomenon whereby people extrapolate beyond the edges of a given view. This revealed that hippocampal activity tracked the emergence of boundary extension, suggesting that scene construction can be rapid, automatic, and implicit.

Overall, my findings shed new light on the nature of episodic representations within the human hippocampus, and offer an empirical link between episodic memory and computational theory. Moreover, they provide further evidence regarding scene construction, which is a key component of episodic representations within the hippocampus.

Acknowledgements

There are a great many people who I have to thank for their contributions to this thesis. First and foremost, thanks to Eleanor Maguire for taking me under her wing, and showing me how to do science properly. I have learned a huge amount during the last three (and a bit) years, and I attribute a lot of this to the amount of time and effort that Eleanor has contributed to each of my projects, both the successes and the failures. I would also like to thank Geraint Rees for being an exemplary second supervisor, and for his constant enthusiasm.

I also have to thank all of my collaborators, and in particular: Demis Hassabis for introducing me to MVPA, and for guiding me through the first few months of my PhD; Heidi Bonnici for teaching me how to segment the hippocampal subfields, and collaborating on so many interesting projects; Sinead Mullally for her work on Boundary Extension, and her general life and neuroscience wisdom. On the technical side I have many people to thank, as none of these projects would have been possible without their advice and support, so thanks to Nikolaus Weiskopf and the physics group, Karl Friston and the methods group, David Bradbury and all of those from imaging support, Ric Davis and the IT department, and to Peter Aston, Marcia Bennet, Marina Anderson, Tom Simpson, and Alison Ryan for guiding me through every variety of admin.

I also owe a huge thanks to my family for accepting my “lifestyle choice” entailed by taking on a PhD, and for supporting me at every step. Most important of all, I owe a huge debt to my loving wife, Cat Sebastian, who has been so supportive of every decision I’ve taken – even my decision to do a PhD despite all the warnings! And most of all, thanks for never once saying “I told you so” when writing this thesis got tough.

Contents

1 Chapter 1 12
1.1 Introduction 13
1.2 Episodic memory 15
1.2.1 What is episodic memory? 15
1.2.2 Do animals have episodic memory? 16
1.3 The neural basis of episodic memory 17
1.3.1 The hippocampus 17
1.3.2 Neuroimaging data 20
1.4 Consolidation of episodic memories 22
1.5 The spatial role of the hippocampus 30
1.6 Scene construction in the hippocampus 35
1.7 Information processing in the brain 41
1.7.1 Neural representations 41
1.7.2 Neural computations 42
1.7.3 Neural processes 43
1.7.4 Memory traces 44
1.8 How are episodic memories represented? 45
1.8.1 What do we know so far? 45
1.8.2 Computational theories of episodic memory 47
1.8.3 Empirical support for the computational theories 52
1.8.4 What does this tell us about episodic memory? 58
1.9 Multi-voxel pattern analysis 59
1.10 Thesis overview 64
1.11 Publications 68
2 Chapter 2 70
2.1 Methods overview 71
2.2 Participants 71
2.3 Experimental Tasks 72
2.4 The biophysics of MRI 72
2.4.1 MR signal generation 72
2.4.2 MR image formation 75
2.4.3 MR scan types 76
2.4.4 The BOLD signal 77
2.4.5 Resolution of fMRI 79
2.5 Specific MRI details 80
2.5.1 MRI scanners 80

2.5.2 MRI sequences 80
2.6 Univariate analysis of fMRI data 83
2.6.1 Analysis overview 83
2.6.2 Spatial preprocessing 85
2.6.3 Mass-univariate statistical analysis 88
2.7 Multi-voxel pattern analysis 94
2.7.1 MVPA methods 96
2.7.2 MVPA classification overview 96
2.7.3 Initial selection of voxels 99
2.7.4 Selecting an MVPA method 102
2.7.5 Data preprocessing 107
2.7.6 Feature selection 109
2.7.7 MVPA versus fMRI adaptation 110
2.7.8 High-resolution fMRI and MVPA 111
2.7.9 The interpretation of classifier accuracies 112
2.7.10 Decoding different levels of information 113
2.8 Segmentation of regions of interest 115
2.9 Dynamic Causal Modelling 116
3 Chapter 3 118
3.1 Introduction 119
3.2 Methods 122
3.2.1 Participants 122
3.2.2 Pre-scan training 122
3.2.3 Task 123
3.2.4 Image acquisition 126
3.2.5 Univariate analysis 127
3.2.6 Image pre-processing for multivariate analysis 128
3.2.7 MVPA classification 129
3.2.8 Feature selection 133
3.2.9 Information maps 136
3.2.10 Overlap analysis 136
3.2.11 Temporal dependencies: control analysis 137
3.3 Results 138
3.3.1 Behavioural Results 138
3.3.2 Univariate Results 140
3.3.3 MVPA Results 140
3.3.4 Information maps 143
3.3.5 Temporal dependencies: control analysis 144

3.3.6 Comparison of cued and free recall conditions 145
3.4 Discussion 145
3.5 Clinical applications 149
4 Chapter 4 152
4.1 Introduction 154
4.2 Methods 158
4.2.1 Participants 158
4.2.2 Autobiographical Memories 158
4.2.3 Pre-scan training 158
4.2.4 Task 159
4.2.5 Image acquisition 160
4.2.6 ROI segmentation 162
4.2.7 Image pre-processing for MVPA analysis 164
4.2.8 MVPA classification 164
4.2.9 Information maps 165
4.2.10 Statistical analysis 166
4.3 Results 166
4.3.1 Behavioural Results 166
4.3.2 MVPA analysis 167
4.3.3 Spatial distribution of information within the hippocampus 170
4.4 Discussion 173
5 Chapter 5 182
5.1 Introduction 183
5.2 Methods 185
5.2.1 Participants 185
5.2.2 Stimuli 185
5.2.3 Pre-scan training 187
5.2.4 Scanning task 188
5.2.5 Post-scan debrief 189
5.2.6 Image acquisition 190
5.2.7 Image preprocessing 190
5.2.8 MVPA analyses 191
5.2.9 Misclassification analysis 193
5.2.10 Permutation testing 194
5.2.11 Controlling for the number of voxels 196
5.2.12 Examining the effects of smoothing 196
5.3 Results 197
5.3.1 Behavioural results 197

5.3.2 Four-class MVPA classification 199
5.3.3 Spatial context MVPA classification 200
5.3.4 Event content MVPA classification 202
5.3.5 Misclassification analysis 202
5.4 Discussion 203
6 Chapter 6 210
6.1 Introduction 211
6.2 Methods 214
6.2.1 Experimental design 214
6.2.2 Image acquisition 214
6.2.3 Data preprocessing 215
6.2.4 Segmentation of the hippocampal subfields 215
6.2.5 Behavioural variables 222
6.2.6 Decoding Analyses 223
6.2.7 Mediation Analysis 229
6.3 Results 230
6.3.1 Distinct episodic information 230
6.3.2 Episodic pattern completion 231
6.3.3 Individual differences in episodic representation 233
6.3.4 Mediation analysis 236
6.4 Discussion 237
7 Chapter 7 244
7.1 Introduction 246
7.2 Methods 251
7.2.1 Participants 251
7.2.2 Procedure 252
7.2.3 Boundary extension task 252
7.2.4 Behavioural analysis 253
7.2.5 Anatomical regions of interest 254
7.2.6 MRI acquisition 255
7.2.7 Image pre-processing 256
7.2.8 Neuroimaging analysis 256
7.2.9 ROI-based analyses 258
7.2.10 Dynamic causal modelling 258
7.3 Results 260
7.3.1 Behavioural results 260
7.3.2 Whole-brain fMRI results 260
7.3.3 ROI analysis 262

7.3.4 Hippocampus– PHC connectivity 264
7.3.5 fMRI adaptation 265
7.3.6 Top-down modulation of visual cortex – DCM results 268
7.4 Discussion 271
8 Chapter 8 277
8.1 Can we detect individual episodic memory representations in the human hippocampus? 279
8.2 How do episodic memory representations change over time? 282
8.3 What is the nature of episodic memory representations in the hippocampus? 286
8.4 How do hippocampal subfields contribute to episodic memory representations? 289
8.5 What is the role of the hippocampus in boundary extension? 292
8.6 Conclusions and future directions 295
8.6.1 How do memory traces evolve over time? 296
8.6.2 What functional dissociations account for the differences in episodic memory representations along the anterior-posterior axis of the hippocampus? 297
8.6.3 What are the precise roles of the hippocampal subfields in episodic memory? 299
8.6.4 What roles do the hippocampal subfields play in individual differences in episodic memory? 300
8.6.5 How do scene construction and boundary extension contribute to the representation of episodic memories at the neural level? 302
8.6.6 Final conclusions 303
9 References 304

List of Figures

Figure 1. The anatomical location and connectivity of the human hippocampus 19

Figure 2. The autobiographical memory network 21

Figure 3. Two theories of the hippocampal-cortical interactions involved in the representation of episodic memory over time 25

Figure 4. Place cell and grid cell activity patterns 32

Figure 5. Amnesic patients show attenuation of boundary extension 39

Figure 6. The subfields of the hippocampus 50

Figure 7. Connectivity of the hippocampal subfields 52

Figure 8. The principles of multi-voxel pattern analysis 61

Figure 9. Magnetic spin 74

Figure 10. Canonical haemodynamic response function 78

Figure 11. MVPA classification example 98

Figure 12. Experimental design 124

Figure 13. The overall MVPA classification procedure 132

Figure 14. The searchlight feature selection procedure 135

Figure 15. MVPA decoding results 141

Figure 16. Hippocampal information maps 142

Figure 17. Information heatmaps 143

Figure 18. Episodic memory decoding in patients with unilateral hippocampal sclerosis and intractable epilepsy 150

Figure 19. Example of timeline from a trial during scanning 160

Figure 20. The brain regions examined 163

Figure 21. MVPA results for recent and remote autobiographical memories 168

Figure 22. Information maps in the hippocampus 171

Figure 23. MVPA results for anterior and posterior subregions of the hippocampus 172

Figure 24. The movies 186

Figure 25. Timeline of a sample trial during fMRI scanning 187

Figure 26. An overview of the decoding analyses 195

Figure 27. Summary of MVPA results 201

Figure 28. Subfield segmentation 220

Figure 29. The movies 225

Figure 30. Episodic information in the hippocampal subfields 232

Figure 31. Correlations with individual differences in awareness of commonalities 235

Figure 32. Mediation Analysis 236

Figure 33. The two phases of boundary extension 249

Figure 34. Example of a single experimental trial 254

Figure 35. Anatomical regions of interest 255

Figure 36. Neural correlates of the boundary extension effect 261

Figure 37. Time-course of the boundary extension effect 263

Figure 38. Modelling hippocampal-PHC connectivity during BE 265

Figure 39. Adaptation effects in early visual cortex reflects changes in subjective perception 267

Figure 40. Modelling hippocampal-visual cortex connectivity 270

List of Tables

Table 1. Free Recall statistical dependencies 127

Table 2. Behavioural results 138

Table 3. Debriefing questionnaire results 139

Table 4. Memory characteristics 167

Table 5. Memory debrief ratings 198

1 Chapter 1

Introduction

1.1 Introduction

Episodic memory, the memory for our personal past experiences, is fundamental to normal human existence. We take for granted the rich tapestry of memories stretching from the present day right back to early childhood which gives each of us a sense of continuity, and forms a core part of our self-identity (Conway and Pleydell-Pearce, 2000). But when this ability is lost, as in amnesia, the result is a debilitating impairment.

The study of the biological basis of episodic memory has a long history, and much evidence points to the hippocampus as a critical neural locus of episodic memory in the human brain (Scoville and Milner, 1957; Maguire, 2001; Spiers et al., 2001; Cipolotti and Bird, 2006; Svoboda et al., 2006). However, we still have a poor understanding of exactly how episodic memories are represented in terms of the underlying neuronal populations within the hippocampus, despite the existence of a number of detailed theoretical models of episodic memory (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). This is largely because episodic memory can only be studied with certainty in humans (Tulving, 2002; Suddendorf and Busby, 2003), which precludes the use of direct animal models which have proved so important to the understanding of other forms of memory such as spatial memory (Andersen et al., 2006).

The core of my thesis is an investigation of episodic representations within the human hippocampus. The majority of my work involves the development and use of multi-voxel pattern analysis (MVPA) or 'decoding'

of functional magnetic resonance imaging (fMRI) data. During the course of my thesis I hope to demonstrate that this approach can provide a useful new perspective on our understanding of the neurobiological substrates of episodic memory.

In this chapter I will define episodic memory, and summarise the current state of knowledge regarding its neural basis. Following this, I will consider one of the major debates within the episodic memory literature – whether consolidation processes render remote episodic memories fully independent of the hippocampus. Although the spatial role of the hippocampus is not the main focus of this thesis, I nevertheless provide a brief overview of this well-characterised function in both rodents and humans as it resonates with some aspects of my work. I then describe recent experiments which demonstrate that the hippocampus plays a critical role in imagining novel events and scenes. In this section I introduce the concept of 'scene construction', which has particular relevance for my final experiment. I then discuss the neural representation of episodic memories in the hippocampus, including hippocampal anatomy and theoretical models. I conclude this section by noting the empirical gap between computational models of hippocampal function and episodic memory. Thus, there is a dearth of concrete knowledge of the underlying neuronal representation of an episodic memory. Following this, I introduce the concept of MVPA as applied to fMRI data, and argue that this approach could offer a means of investigating information at the level of specific episodic memory representations within the human hippocampus. Finally, I provide an overview of the thesis including the specific aims of my research studies.

1.2 Episodic memory

1.2.1 What is episodic memory?

The concept of episodic memory as a distinct type of memory was explicated by Endel Tulving in 1972 (Tulving, 1972), who defined it as the memory for personally experienced events. This definition therefore differentiated episodic from semantic memory, which was proposed to be abstracted, general knowledge about the world. While the current concept of episodic memory overlaps considerably with the original, it has evolved in the four decades since the first seminal publication (Tulving, 2002). Episodic memory is now defined as the memory for personally experienced events in our lives which includes both information about the content of the event, and the specific spatial and temporal context of that event (the "what, where, and when" of episodic memory – Tulving, 1983). In addition to this specific memory content, true episodic recollection entails a rich re-experiencing of the past event. Episodic memory as a concept is therefore based on both the content of the memory (what, where, and when), and the conscious experience of the retrieved memory, and both of these factors are required for a memory to be considered genuinely episodic. This stands in stark contrast to semantic memory, which refers to factual knowledge about the world which is usually acontextual, and is not accompanied by vivid recollective experience. Usually semantic memory refers to abstracted knowledge such as concepts, words, categories, but also includes autobiographical knowledge about ourselves, such as our name and where we live. These latter types of memory are sometimes referred to as personal semantics (Kopelman et al., 1989), which are related to, but distinct from

episodic memory. It is even possible to have semantic knowledge of a specific event, including abstracted knowledge about the location of that event and the temporal context. Thus, on occasion it may be possible to have a fully semanticized memory of what, where, and when, but with no accompanying rich re-experiencing of the event. This latter example demonstrates the necessity of defining episodic memory in terms of both the content and the phenomenology of the retrieval experience. I will be using this definition of episodic memory throughout the thesis, and in each study I placed a particular emphasis on the vivid recall of naturalistic episodes in order to ensure that the memories studied were truly episodic in nature. That is not to say that this is the only possible definition of episodic memory, as some might argue that the content alone (what, where, when) should define episodic memory. Nevertheless, the operational definition I will use throughout this thesis encompasses both the content and the phenomenal experience of episodic recall. Note that one important sub-type of episodic memory is autobiographical memory (Conway and Pleydell-Pearce, 2000), which describes an episodic memory for personally meaningful experiences. I will be investigating both episodic and autobiographical memories in the course of this thesis.

1.2.2 Do animals have episodic memory?

Tulving has argued that episodic memory evolved only recently, and is probably unique to humans (Tulving, 2002). This is, however, a controversial argument, as it has been demonstrated that at least some other species show a striking ability to remember “what, where, and when” (Clayton et al., 2003; Eacott and Easton, 2010; Salwiczek et al., 2010).

Despite these impressive abilities to remember the spatio-temporal context of events, it is not possible to determine whether other species have the same rich conscious experience during episodic retrieval, which is the second important criterion of genuine episodic memory (Tulving, 2002). It is currently only possible to assess the phenomenology of memory retrieval through verbal communication, thereby ruling out all other species *a priori*. While it may, in the future, become possible to produce irrefutable evidence that other species do indeed experience episodic memories in the same way as us, for now we can only study true episodic memory in humans (Suddendorf and Busby, 2003). That is not to say, of course, that we cannot learn a great deal about critical neural components underlying episodic memory (such as the "what, where, and when") from studying animals, but for studying vivid recollective experience, we are limited to humans. In the next section, I summarise the current state of knowledge regarding the biological basis of episodic memory in humans.

1.3 The neural basis of episodic memory

1.3.1 The hippocampus

It has been clear for many decades that a structure within the medial temporal lobe (MTL) called the hippocampus (see Figure 1) is critical for the encoding and retrieval of episodic memories (although there is debate over whether the retrieval role of the hippocampus in episodic memory is time-limited – see section 1.4 for a discussion of this issue). The first evidence for this came from the tragic case of Henry Gustav Molaison, better known as patient HM. He suffered from intractable temporal lobe

epilepsy, and elected to undergo an experimental surgical intervention, which involved bilateral resection of his MTL. While the surgery was effective in treating the epilepsy, it also rendered HM densely amnesic (Scoville and Milner, 1957). His impairments were selective to memory function, and did not produce any obvious loss of other cognitive function such as executive function, language, or perception. Nevertheless, he was rendered unable to recall any details of his day-to-day life, could not find his way around, and failed to recognise people that he saw even on a frequent basis (Corkin, 2002). While HM's lesions also included other parts of the MTL, there have been many cases since then which have involved bilateral lesions that seem limited to the hippocampi (as far as can be determined by the use of current MRI technology) which have also produced severe episodic memory impairments (Spiers et al., 2001; Cipolotti and Bird, 2006), demonstrating that the hippocampus itself is critically important to this type of memory. Throughout this thesis I will define the hippocampus as consisting of the CA fields, the dentate gyrus, and the subiculum.

It is important to note that amnesia can be divided into two types – anterograde and retrograde. Anterograde amnesia refers to the inability to form new memories after the occurrence of the hippocampal damage, while retrograde amnesia refers to the loss of memory for events prior to the damage. Amnesics generally have both forms of amnesia to some degree, but the exact gradient of retrograde episodic impairment in amnesia is currently a matter for debate (see section 1.4).

Overall the lesion data is unequivocal in its support of the conclusion that

the hippocampus is vital to episodic memory. This conclusion is further supported by evidence from neuroimaging results, which I describe in the next section.

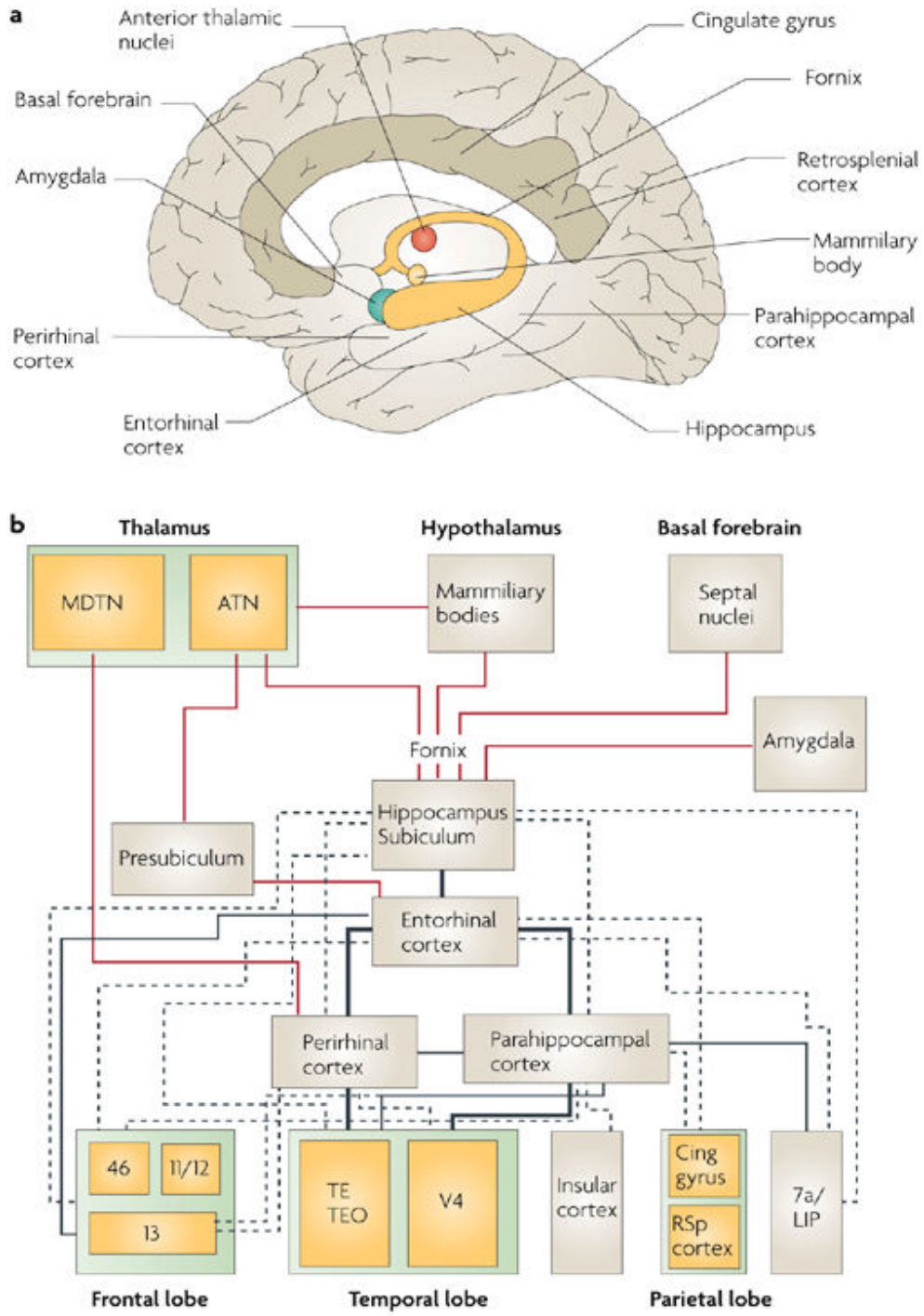

Figure 1. The anatomical location and connectivity of the human hippocampus. *(a) This panel displays the hippocampus in orange in a sagittal cut through the human brain. Also displayed in orange is the fornix,*

which projects posteriorly and superiorly, connecting the hippocampus to other cortical and subcortical structures such as the anterior thalamic nuclei (ATN), the mammillary bodies and the retrosplenial cortex. Immediately neighbouring the hippocampus are the cortical components of the medial temporal lobe, which can be divided into the entorhinal, perirhinal, and parahippocampal cortices. The amygdala is depicted in green and sits immediately in front of, and on top of the anterior hippocampus. (b) This diagram depicts the complex set of connections to and from the hippocampus. The subcortical connections are indicated by the red lines, and the cortical connections by the black lines. The thickness of each line indicates the relative strength of connection between those regions. The majority of cortical inputs to the hippocampus come from the perirhinal and parahippocampal cortex, through the entorhinal cortex. The cortical output of the hippocampus comes from the hippocampal subiculum, which projects back into entorhinal cortex. From Bird and Burgess (2008) with permission from Nature Publishing Group.

1.3.2 Neuroimaging data

In the last two decades, the advent of functional neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have allowed the in vivo investigation of neural activity across the healthy human brain in a non-invasive fashion. These advances have produced a revolution in the way that human cognition can be studied, and this includes episodic memory. Functional neuroimaging studies have consistently found that core network of regions that co-activate during the retrieval of episodic memories (Maguire, 2001; Svoboda et al., 2006). In addition to the hippocampus, these include medial and lateral prefrontal cortex, parahippocampal cortex, lateral temporal regions retrosplenial cortex, temporo-parietal junction, thalamus, posterior cingulate cortex, and the cerebellum (see Figure 2). It is clear, therefore, that the retrieval of episodic memory involves a widely distributed network of activation.

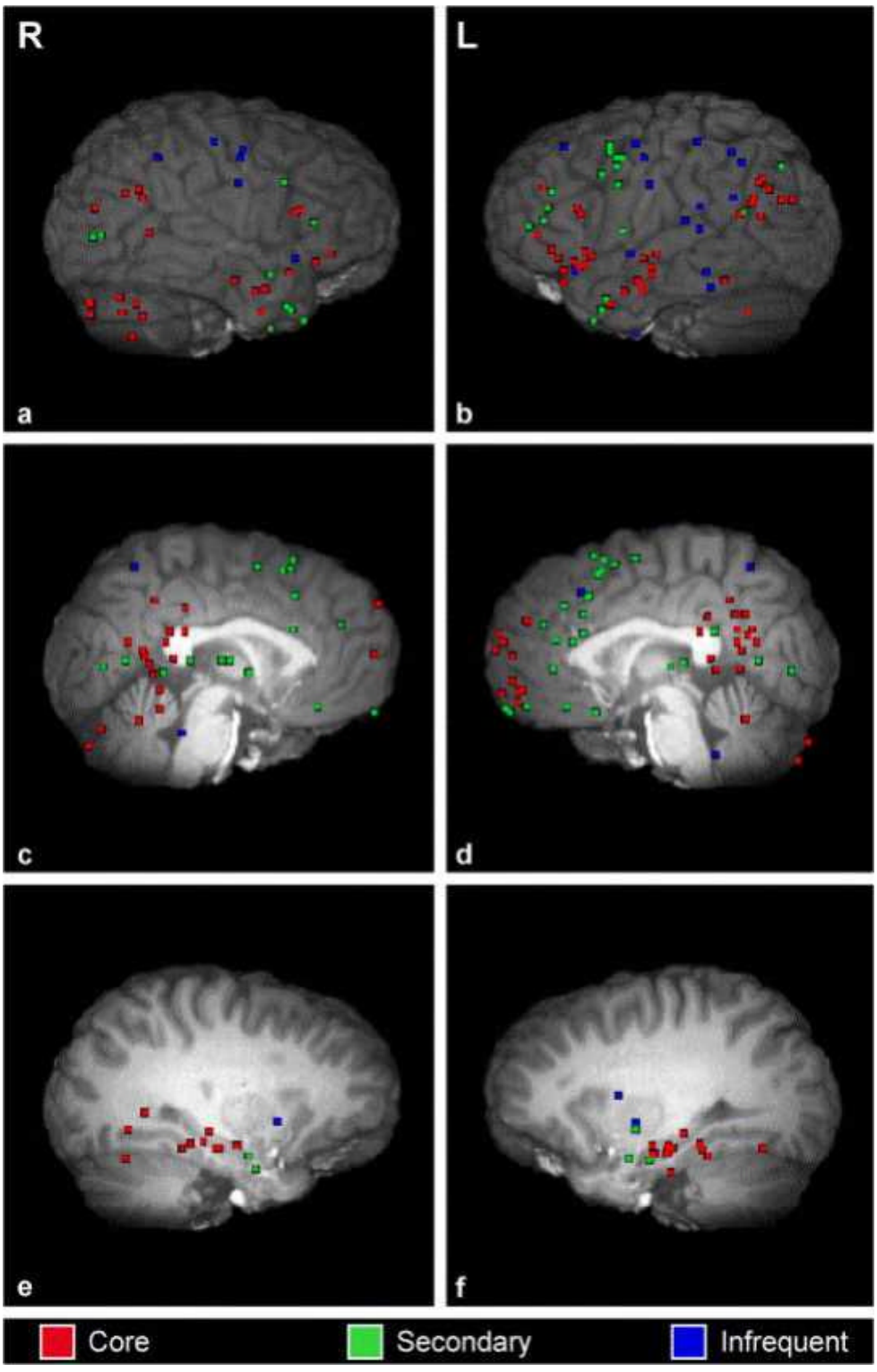

Figure 2. The autobiographical memory network. *This figure is taken from Svoboda et al. (2006), who conducted a meta-analysis on the functional neuroimaging studies of autobiographical memory (AM). Each coloured marker indicates a peak activation from one of the reviewed studies, and these are divided into three categories. The regions which show common activation across AM studies are labelled in red, and these constitute the core AM network, which include the hippocampus, parahippocampal cortex, retrosplenial cortex, posterior cingulate cortex, medial frontal cortex, lat-*

eral temporal and parietal regions, and the cerebellum. Regions which are less frequently activated are designated as secondary (in green) and infrequent (in blue). Reproduced with permission from Elsevier.

While there are numerous theories regarding the particular role of each of these cortical regions within episodic memory (e.g. Wheeler et al., 1997; Rugg et al., 2002; Simons and Spiers, 2003; Cabeza et al., 2008), none have yet been able to convincingly account for all empirical data. Despite the disagreements over the specific contribution of each region to episodic memory, there is wide consensus that each of these regions does play an important role. Nevertheless, it is important to note that damage to any of these cortical regions fails to produce the same kind of severe memory deficits seen after hippocampal damage (with the possible exception of the retrosplenial cortex, which is heavily interconnected with the hippocampus – see Vann et al., 2009). This suggests that it is the hippocampus itself which forms the core structure within this distributed episodic network. While it will undoubtedly be important to develop a better understanding of the contribution of each cortical region to episodic memory and their functional connectivity, the focus of this thesis will be on the hippocampus itself.

1.4 Consolidation of episodic memories

A key concept within the neurobiology of memory is that memories are not completely formed at the instant of encoding, but instead take some time after the initial encoding event to become consolidated into stable neural memory traces (Marr, 1971; Frey and Morris, 1998; Andersen et al., 2006; Redondo and Morris, 2011). This can be broadly separated into (a) cellular

consolidation, which involves the stabilisation of the memory trace at the local, synaptic level, which takes place within minutes or hours, and (b) systems consolidation, which involves the reorganisation of information at the level of entire neuronal networks, and has been described as occurring within the timescale of hours to decades (Winocur et al., 2010; Winocur and Moscovitch, 2011). Cellular consolidation is a well-characterised phenomenon, including processes such as long-term potentiation (Bliss and Collingridge, 1993) and synaptic tagging (Frey and Morris, 1998; Redondo and Morris, 2011). Systems consolidation, on the other hand, is much less well characterised due to the complexity of studying the dynamic interaction of multiple neural systems over time. Early neurobiological theories proposed that the interaction between the hippocampus and neocortex is a classic example of systems-level consolidation. For instance, David Marr (1971) conceptualised the hippocampus as a time-limited memory store, after which the memory would be consolidated into the neocortex (for more details of this theory, see section 1.8.2). Following this consolidation period, the hippocampus is no longer necessary for the retrieval of that memory (see Figure 3, top panel).

There is clear evidence that systems-level consolidation can occur in animals, as retrieval of certain types of information (e.g. object recognition or contextual fear conditioning) only depend on the hippocampus for a limited period of time (Winocur, 1990; Zola-Morgan and Squire, 1990; Kim and Fanselow, 1992; Takehara et al., 2003). Indeed, recent evidence demonstrates that such consolidation can occur remarkably quickly, depending on the degree of previous knowledge or “schemas” (Tse et al.,

2007, 2011). At the same time, however, it has also been demonstrated that at least some types of memory, such as allocentric spatial memory, do not show any kind of temporal gradient, even when tested nine months after learning (Sutherland et al., 2001; Clark et al., 2005a, 2005b; Winocur et al., 2005). Thus, the data from animal studies of hippocampal consolidation are mixed, and suggest that some types of memory may always depend on the hippocampus (Winocur et al., 2010).

Given this equivocal set of results from the animal literature, one obvious question is whether episodic memories are subject to a consolidation process, or whether episodic retrieval always depends on the hippocampus. Until about 15 years ago, the prevailing view was that episodic memory is indeed subject to a consolidation process, and this view was championed by Larry Squire and his colleagues (Squire, 1992; Squire and Alvarez, 1995; Squire and Zola, 1998; Squire et al., 2004). This account grew out of neuropsychological studies of patients with MTL damage, who showed a temporal gradient in retrograde amnesia (Marslen-Wilson and Teuber, 1975; Rempel-Clower et al., 1996; Squire and Bayley, 2007). In other words, these patients showed serious deficits in recalling recent memories which had not yet been consolidated into the neocortex. More remote memories, on the other hand, were relatively intact, because these older memories had been fully consolidated. This view is now widely described as the Standard Consolidation Theory, or Standard Model of Consolidation, and I will use both of these terms in this thesis.

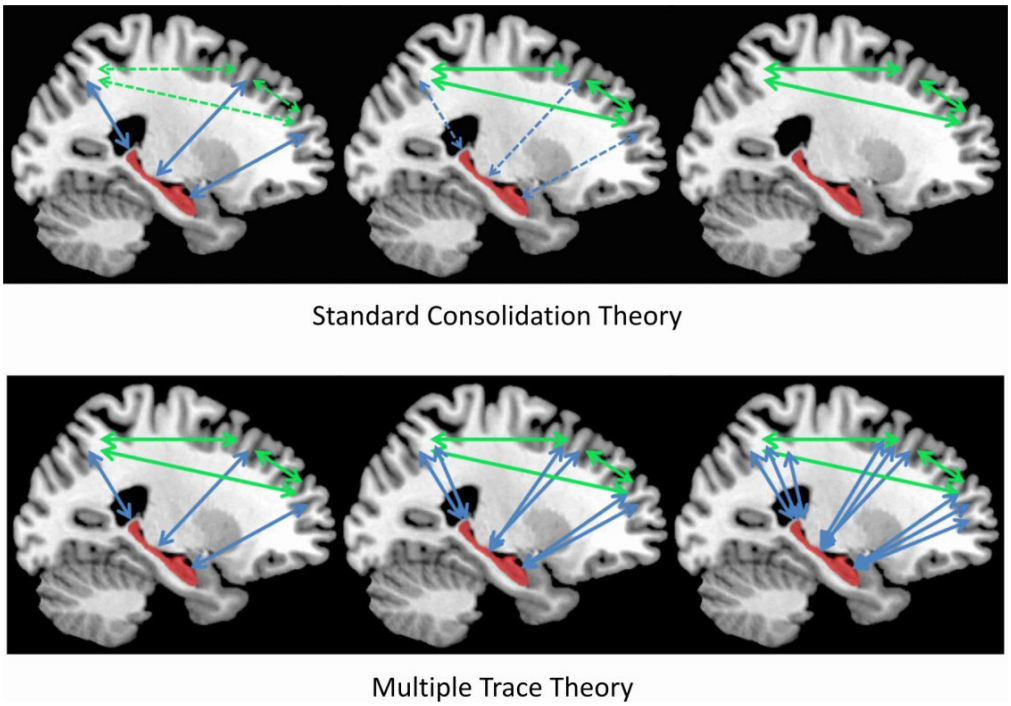

Figure 3. Two theories of the hippocampal-cortical interactions involved in the representation of episodic memory over time. *The top panel displays the consolidation process proposed by the Standard Consolidation Theory. This theory proposes that a newly encoded episodic memory is widely distributed throughout the cortex, but depends on connections with the hippocampus for retrieval. Over time however (left to right), the memory consolidates into the neocortex as the connections between the cortical regions strengthen. As this happens, the memory becomes less dependent on the hippocampus until eventually the memory is fully consolidated, and completely independent of the hippocampus. This process is proposed to happen very slowly over a number of years, or even decades. The bottom panel displays the hippocampal-cortical representation of episodic memory over time as proposed by the Multiple Trace Theory. According to this theory, a newly encoded episodic memory is widely distributed throughout the cortex, and depends on connections with the hippocampus for retrieval. Each time a memory is retrieved (left to right), a new hippocampal memory trace is created, which strengthens the hippocampal-cortical connections for that memory. The creation of these multiple traces facilitates the extraction of semantic information from the episode, which is stored within the cortex. Eventually, this can lead to the creation of a fully semanticized memory, which can exist independently of the hippocampus. Thus, according to this account, genuinely episodic recall always depends on the hippocampus, no matter how old the memory, while semantic memories can show a temporal gradient, becoming more cortically dependent over time.*

This view was challenged by a paper by Lynn Nadel and Morris Moscovitch (1997), who pointed out that there are cases of retrograde amnesia that do not show any evidence for a temporal gradient in episodic memories. Indeed, even those patients who do show a temporal gradient only do so for memories which are more than 25 years old. They argued that a consolidation process occurring over this kind of time-span cannot be a biologically adaptive process, as suggested by the Standard Consolidation Theory. Out of this observation (along with other problems with the standard consolidation theory), these authors proposed an alternative model of consolidation which they termed the Multiple Trace Theory (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Winocur et al., 2010; Winocur and Moscovitch, 2011). According to this theory, episodic memory traces are not consolidated "out" of the hippocampus into neocortex. Instead, each time a memory is retrieved, a new hippocampal memory trace is created, and each of these traces shares some of all of the information about the initial episode (see Figure 3, bottom panel). The creation of multiple, related traces facilitates the extraction of semantic information from the episode, which is stored in the cortex. Over time, and with enough repeated retrieval, this can lead to the creation of a semanticized or schematic memory for a memory which can be hippocampally independent. Truly episodic retrieval (the rich re-experiencing of an event, including its spatiotemporal context, consistent with Tulving's definition), however, always requires the hippocampus. According to this model, the long-term temporal gradient sometimes found in retrograde amnesia reflects the gradual semanticization of episodic memories, and it is proposed that these patients can never retrieve genuinely episodic memories.

The neuropsychological literature is somewhat divided on this issue, as some patients have been shown to have a temporal gradient in episodic memory (Squire, 1992; Squire et al., 2004; Squire and Bayley, 2007) while others have not (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Winocur et al., 2010; Winocur and Moscovitch, 2011), and a recent review of the literature concluded that there were roughly equal numbers of patients in each of these groups (Winocur and Moscovitch, 2011). How then, do the two different theories reconcile these different results? Squire consistently argues that patients who do not show a temporal gradient effect must have damage that extends beyond the hippocampus (Squire et al., 2004). This argument is in fact impossible to rule out, as it would involve proving a negative - that there is no damage outside the hippocampus in any of the patients. Given that it could always be argued that there are subtle types of damage beyond that which is measurable with current techniques, it is not feasible to investigate this. Thus, while this is a logically plausible argument, it requires the explicit demonstration that those patients that do not show a temporal gradient of retrograde amnesia have damage above and beyond the hippocampus, and furthermore, to characterise exactly what kind of damage leads to this effect.

A comparison of patients with hippocampal damage to those with hippocampal damage plus damage to cortical regions outside of the MTL (MTL+) was conducted by Bayley et al. (2005) in order to try and provide this kind of evidence. They show that patients with damage restricted to the hippocampus show spared remote episodic memories, while those with additional cortical damage do not. Notably, however, the extra-hippocampal

damage was highly variable across the MTL+ patients, which makes it much more difficult to come up with a specific hypothesis about what kind of damage would be expected to cause impairments to remote memories. Furthermore, a study by Rosenbaum et al. (2008) found that the extent of damage to the hippocampus itself, rather than neocortex, predicted the extent of retrograde amnesia. Thus, the evidence one way or another is currently inconclusive.

How does the Multiple Trace Theory explain the mixed set of data? Moscovitch and colleagues have a clear hypothesis regarding the existence of preserved remote memories, and this was an explicit part of the original hypothesis (Nadel and Moscovitch, 1997). They argue that, with repeated retrievals, more semantic information about that particular event is extracted, and stored within cortical regions. Eventually, this can lead to the "semanticization" of an episodic memory, such that a large amount of specific information can be retrieved about that memory, but in the absence of true episodic recall. Thus, if the hippocampus is damaged, then these highly semanticized episodes can still be retrieved, but the quality of the memory will be distinct from true episodic memories. Specifically, details regarding the spatiotemporal context will be impaired, and the memory will not be vividly re-experienced. However, evidence for this assertion is currently mixed, with some studies providing potential support (Maguire et al., 2006; Hassabis et al., 2007a Rosenbaum et al., 2008), and others not (Bayley and Squire, 2005; Kirwan et al., 2008). Thus, to date, the neuropsychological evidence does not clearly support either the Standard Consolidation Theory or the Multiple Trace Theory, and no completely

convincing explanation has yet been found to account for the discrepancies between the different neuropsychological studies.

The evidence from functional neuroimaging is more clear-cut, as the majority of studies have found that the hippocampus is active for the recall of both recent and remote memories (for reviews see Maguire, 2001; Svoboda et al., 2006). Squire et al (2004) have argued that this kind of activity could simply reflect activity related to the encoding of the new information within the scanner. However, even when this is specifically controlled for, the hippocampus still displays activation for the retrieval of remote episodic memories (Gilboa et al., 2004). Interestingly, both this study and another by Addis et al. (2004) have found that the activity within the hippocampus correlates with the richness and vividness of the recalled memory. Taken together, these results suggest that the hippocampus is always involved in the retrieval of episodic memory so long as those memories are vividly re-experience in rich detail, which is consistent with the Multiple Trace Theory.

Indeed, Moscovitch et al. (2005) argued that the discrepancies between the various patient studies may be attributable to similar differences in the quality of episodic recall, suggesting that seemingly spared remote memories were based on a more semanticized representations. However, studies using more sensitive measures of episodic memory that purportedly allow the separation of episodic and semantic details (e.g. the Autobiographical Interview – Levine et al., 2002) have still found mixed results (Steinvorth et al., 2005; Kirwan et al., 2008; Rosenbaum et al., 2008),

with no clear evidence for this assertion.

Overall, the picture is still not clear regarding episodic consolidation. The neuroimaging evidence clearly points to a role for the hippocampus in episodic retrieval regardless of the age of memories, but it could be argued that this evidence is correlational, and cannot tell us whether or not the hippocampus is actually necessary for the retrieval of remote memories (Squire et al., 2004). Thus, the lesion studies are critical for establishing this point, and the results of these studies are more mixed. In order to break this theoretical dead-lock, it may be necessary to find innovative new approaches to investigate the hippocampal role in the representation of remote episodic memories, which is a goal of my thesis.

1.5 The spatial role of the hippocampus

While episodic memory has been the major focus of human hippocampal research for the last few decades, rodent research has tended to focus on a second well-documented function of the hippocampus – spatial representation. O'Keefe and Dostrovsky (1971) were the first to report the extraordinary firing characteristics of "place cells" in the rat hippocampus (Figure 4). These neurons fire when an animal is in a specific location in an environment, regardless of the direction where the animal is headed or where it is looking. A population of place cells is therefore potentially able to represent an entire spatial environment in terms of discrete allocentric locations, which has been dubbed a "cognitive map" by O'Keefe and colleagues (O'Keefe and Nadel, 1978; Andersen et al., 2006). Importantly,

the location of place cell firing within a familiar environment (the “place field”) remains stable for several weeks, indicating that place cells may encode a long-term memory for spatial environments (Lever et al., 2002). Evidence for neurons that appear to be similar to place cells has now been documented in monkeys (Ono et al., 1991) and humans (Ekstrom et al., 2003; Hassabis et al., 2009), demonstrating that this may be a core function of the hippocampus that is preserved across different species.

More recently, another important type of spatial neuron has been discovered within the medial entorhinal cortex of the rat (Hafting et al., 2005). These neurons also display spatially selective patterns of activity within an environment, but rather than firing for one location only, they fire for multiple locations within an environment. These multiple spatial locations form a remarkably consistent grid-like pattern across a given environment, and this characteristic has led to these neurons being dubbed “grid cells” (Figure 4). The spatial layout of grid cell firing peaks forms a repeating hexagonal lattice across the environment, and are expressed immediately in a novel environment (Moser et al., 2008). There is some evidence to suggest that the specific spatial distance between the individual points in a grid varies systematically between caudal and ventral grid cells (Moser et al., 2008), potentially allowing the representation of space and distance on many different spatial scales. Grid cells have since been found to exist in fruit bats (Yartsev et al., 2011), and there is now some neuroimaging evidence suggesting that human entorhinal cortex may also contain grid cells (Doeller et al., 2010). There have been several theories proposed to explain the relationship between place cells and grid cells, and how the two

interact to form a complete cognitive map (McNaughton et al., 2006; Burgess et al., 2007; Moser et al., 2008). Altogether, there is now a large body of evidence demonstrating that neurons within the hippocampus and neighbouring post-subiculum, thalamus and retrosplenial cortex (Taube, 2007; Bird and Burgess, 2008; Moser et al., 2008; Vann et al., 2009; Derdikman and Moser, 2010) display highly specialised patterns of activity consistent with the representation of allocentric space, and that this is preserved across different species.

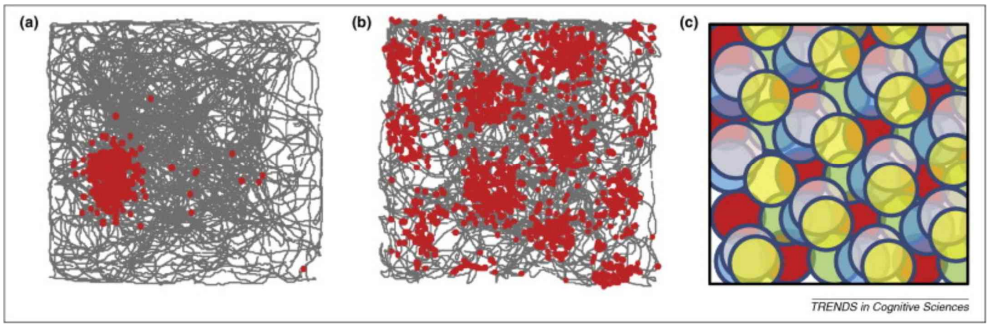

Figure 4. Place cell and grid cell activity patterns. *(a) Example of the activity displayed by a hippocampal place cell in an open-field box. The movement trajectory of a rat is marked by a grey line, and the positions within the environment where the cell fires are marked with a red dot. It is clear that this place cell shows a highly specific preference for one location within the environment. (b) Example of the activity displayed by an entorhinal grid cell. In this case the cell does not have a preference for just one location, but instead shows multiple firing loci, which form the classic hexagonal grid pattern. (c) This diagram represents the spatial activation which would be displayed by a population of grid cells. Here it is clear that just five cells with grids that are slightly offset from one another can comprehensively cover the entire environment. Such an arrangement offers a potentially highly efficient way of coding for any given location within that environment. From Derdikman and Moser (2010) with permission from Elsevier.*

Lesion studies have provided a second important source of evidence for the hippocampal role in spatial memory. Among the most compelling evidence for this was produced by Richard Morris and colleagues (1982) who showed that rats with hippocampal lesions were impaired in a water maze task (this

study was so influential that the paradigm has since been named the Morris water maze). The set-up is very simple – on each trial, a rat is placed in a pool of water, and has to navigate to a platform which is hidden under the surface of the water. They are naturally motivated to do this to avoid the mild stress of constantly having to swim. Rats will ordinarily learn the location of the hidden platform after only a few trials, whereas rats with hippocampal lesions are impaired (Morris et al., 1982). Interestingly, this is only the case when the rats are placed in a different starting location on each trial. If they start at the same location each time, then hippocampal lesions do not lead to navigation deficits, demonstrating that the hippocampus is critical for the allocentric, but not egocentric representation of space (Eichenbaum et al., 1990).

Hippocampal lesions in humans have also been linked to deficits in spatial memory (Burgess et al., 2002). For example, patients with bilateral hippocampal damage show impairments in the ability to learn new environments (Burgess et al., 2002), and to retrieve familiar environments (Maguire et al., 2006). There is also evidence that more basic aspects of spatial memory and perception may be impaired following hippocampal damage (Lee et al., 2005a; Hartley et al., 2007). Consistent with this, functional neuroimaging studies have consistently demonstrated an important role of the hippocampus in spatial navigation and memory (Hartley et al., 2003; Spiers and Maguire, 2006; Viard et al., 2011). Finally, there is evidence that the hippocampi of trainee taxi-drivers display structural (grey matter volume) changes consistent with plasticity during the course of extensive training on the spatial layout of London (Woollett and

Maguire, 2011). Put together, there is substantial support for the proposal that the hippocampus is critical for the representation of allocentric spatial information.

Given the clear role of the hippocampus in both episodic and spatial memory, it has been suggested that episodic memory critically depends on the spatial functions of the hippocampus (O'Keefe and Nadel, 1978; Burgess et al., 2002). The original formulation of this was put forward in the form of the cognitive map theory, which argued that genuine episodic memory is inherently spatial in nature, and that the hippocampus therefore provides the spatial scaffold on which episodic memory is built (O'Keefe and Nadel, 1978). An alternative view of the relationship between spatial and episodic memory is offered by the relational theory, which argues that the primary function of the hippocampus is not spatial, but should instead be thought of as the representation of associations between disparate elements (Cohen and Eichenbaum, 1993; Eichenbaum, 2004). Specifically, this theory posits the existence of three elemental cognitive processes that are all mediated by the hippocampus - associative representation, sequential organisation, and relational networking. According to this view, these fundamental properties can fully account for the spatial processing found within the hippocampus, and are flexible enough to explain the possible non-spatial hippocampal processes. This debate is not the major focus of the thesis, so I will not go into further detail here. The crucial point is that both of these theories agree that the hippocampus is critical for both spatial and episodic memory.

1.6 Scene construction in the hippocampus

While the role of the hippocampus in both episodic and spatial memory has been widely accepted for many years, more recently there have been a number of studies suggesting that the function of the hippocampus may extend beyond these functions. For instance, there is now evidence suggesting that the hippocampus may be involved in short-term and working memory, and even in online perception (Lee et al., 2005a, 2005b; Hannula et al., 2006; Olson et al., 2006; Hartley et al., 2007), although these effects tend to be subtle. More impressive still is the profound deficit found in amnesic patients when asked to imagine a future scenario, or imagine a novel scenario in the present (Klein et al., 2002; Hassabis et al., 2007a; Rosenbaum et al., 2009; Andelman et al., 2010; Race et al., 2011). While healthy control participants have no problem generating and reporting detailed visual scenes, patients with hippocampal damage are severely impaired. This impairment was particularly pronounced on indices of spatial coherence among the elements of a scene, suggesting that the core deficit may be the ability to combine elements into a single, coherent spatial scene or event (Hassabis and Maguire, 2007, 2009; Hassabis et al., 2007a). Functional imaging studies have further confirmed the role of the hippocampus in both imagined future events and current scenes (Okuda et al., 2003; Addis et al., 2007; Hassabis et al., 2007b; Szpunar et al., 2007; Botzung et al., 2008), and together these results clearly demonstrate that the hippocampus is critical for imagination of scenes and events as well as recalling past experiences. These findings have provoked a re-evaluation of

our assumptions regarding the function of the hippocampus, as none of the major theories can immediately accommodate this role in imagination.

Hassabis and Maguire have proposed that the hippocampus may be involved in a core process called ‘scene construction’, which is necessary for imagination and episodic recall (Hassabis and Maguire, 2007, 2009). Scene construction is defined as the process of mentally generating and maintaining a complex and coherent scene or event (and here I define a scene as a view of a real-world, or potential real-world environment comprising background elements and objects arranged in a spatially coherent and appropriate manner. This is the definition I will use throughout the thesis). This necessitates the retrieval and integration of the relevant components of the scene from modality-specific cortex, which are then bound into a coherent spatiotemporal representation. Notably, this concept is flexible enough to account for both newly imagined scenes and retrieved episodic memories, as this core process is held to be involved in both. The authors also argue that scene construction may also be critical for other functions such as spatial navigation and planning for the future (see also Spreng et al., 2009 for a meta-analysis). For example, when trying to remember or decide on a route, we typically visualize scenes from that route in order to aid our memory. This kind of mental simulation requires scene construction in order to conjure up a vivid representation of the spatial environment. This view of episodic memory is consistent with a large body of evidence suggesting that episodic memory is not simply a perfect record of past events, but instead should be considered more of a reconstructive process (Bartlett, 1932; Schacter et al., 1998; Conway and Pleydell-Pearce,

2000). Moreover, scene construction differs from the cognitive map theory in placing the internal representation of scenes at the centre of hippocampal processing, although the hippocampal contribution to scene construction may still, at a fundamental level, be spatial – see more on this below.

Overall, scene construction is an appealing concept, as it brings together various different hippocampal functions and offers an explanatory framework for the shared processes underlying them (for an interesting take on how the spatial processing circuits might contribute to scene construction see Byrne et al., 2007; Bird and Burgess, 2008). However, there are still numerous outstanding questions to address. For instance, what neural processes are involved in scene construction? The processes involved in the retrieval of an episodic memory from a cue have been well-characterised (see section 1.8 for more details on this), but it is not clear how a verbal instruction to imagine a scene can stimulate the hippocampus to generate a completely novel scene representation. Additionally, scene perception has been extensively studied in both the cognitive and neuroscience literature (e.g. Henderson and Hollingworth, 1999; Epstein, 2008), and the parahippocampal cortex has usually been designated the locus of scene perception rather than the hippocampus (but see Mullally and Maguire, 2011, for new insights into why parahippocampal cortex may not in fact be 'scene-selective'). It will therefore be important to determine exactly what the contribution of each of these regions is to both scene perception and scene construction, and how these two processes might relate to one another.

One recent new insight into scene construction comes from a neuropsychological study by Mullally et al. (2012), who studied a cognitive phenomenon known as boundary extension (BE) in a group of amnesic patients. BE is a robust and consistent effect whereby participants remember seeing more of a scene than was present in the physical input, incorrectly extrapolating beyond the physical borders of the original stimulus (Intraub and Richardson, 1989). It is hypothesised that this effect is the result of a two-stage process: when we view a scene, we construct an internal representation of a scene that extends beyond its given border. At retrieval, we then incorrectly believe that the original scene contained more space around the edges, leading to an extension of the boundaries (Intraub, 2012). Mullally and colleagues reasoned that this effect likely requires hippocampal-dependent scene construction processes. In order to test this hypothesis, they assessed the amount of BE displayed by a group of seven amnesic patients who all had selective bilateral hippocampal damage.

Consistent with their prediction, they found that the patients showed significant reductions in BE compared to the healthy control group (Figure 5). This result therefore suggests that the hippocampal role in the representation and construction of scenes may actually go beyond the kind of explicit imagination investigated by Hassabis et al. (2007a), demonstrating that it may also be involved in the automatic, implicit construction of a larger scene when we are simply looking at a picture.

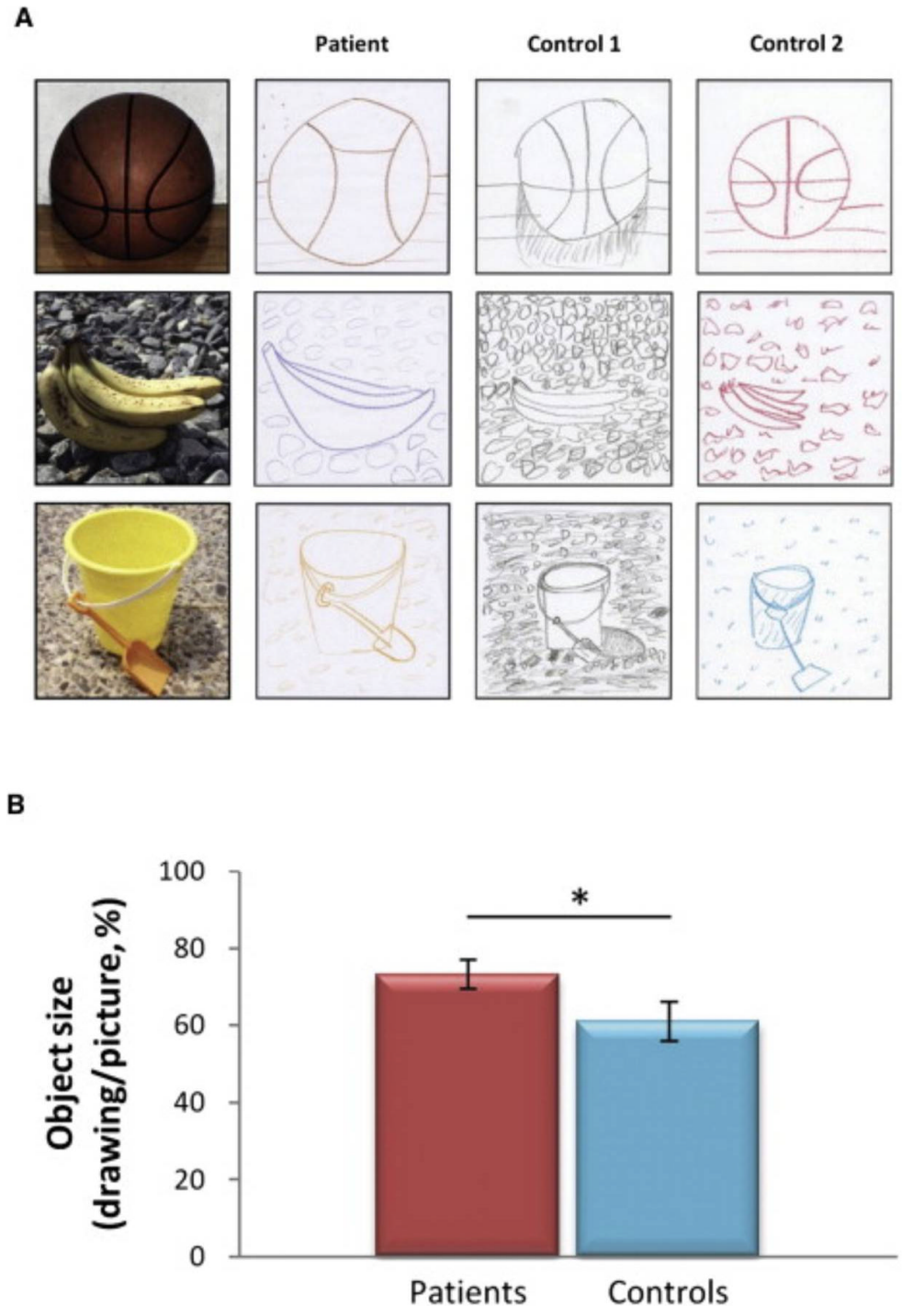

Figure 5. Amnesic patients show attenuation of boundary extension. *(A) In the boundary extension drawing task, participants were asked to study one of the scene pictures on the left for 15 seconds, and then immediately draw it from memory. The two example controls shown here both show a clear boundary extension effect, whereby they incorrectly include more background than was actually present. The example amnesic patient, however, produced a much more accurate depiction of each scene. (B) The boundary extension score was calculated by dividing the size of the original depicted object by the drawn object. This ratio score gives an indication of how much additional space was included around the edges of the object, thereby providing an index of boundary extension (the lower the ratio, the greater the boundary extension effect). As a group, the amnesic patients*

showed significantly less boundary extension than control participants. Notably, this means that the patients were actually more accurate at depicting the original scenes than were the controls, demonstrating that this effect cannot be attributed to a memory error. From Mullally et al. (2012) with permission from Elsevier.

Mullally et al. (2012) devised a further task in order to explore in detail the nature of participants' internal representations of what might be beyond the current view of a scene. This "scene probe" task asked participants to view a close-up view of a scene, and imagine taking a few steps back from the camera's current position, and describing the scene beyond the current view. Patients were able to list appropriate contextual items that might be expected beyond the borders of the scene, and showed no significant difference from controls in this respect. However, the patients provided almost no spatial references regarding the scene beyond the boundaries, and rated the vividness of the extended scenes as significantly lower than controls. Thus, although patients could provide rich semantic and associative information relating to each scene, they could not imagine the spatial structure of the scene. This therefore suggests that the BE deficit seen in this set of patients is also likely to be due to specifically spatial processes rather than any semantic associative processes, which resonates with the original spatial coherence problem uncovered by Hassabis et al (2007a). This intriguing set of findings require further study in order to better characterise the role of the hippocampus in boundary extension, and this is another goal of my thesis.

Overall, there is growing evidence that scene construction is a critical function of the hippocampus, and that this may be one of the core

components involved in episodic recall, spatial navigation, as well as other prospective functions. What remains to be determined is under exactly what circumstances scene construction is required, and how it might interact with other key components of episodic memory, such as the rapid encoding of complex spatiotemporal associations. In the next two sections I will discuss the specific neural mechanisms proposed to underlie these latter aspects of episodic memory.

1.7 Information processing in the brain

So far I have demonstrated that the hippocampus is vital for episodic memory (as well as spatial memory and scene construction). However, in order to develop a complete understanding of episodic memory, we have to consider how episodic memories are physically represented in terms of the underlying neuronal populations. In other words, what does an episodic memory "look like"? Before I turn to this specific question, it will be necessary to define four key concepts for understanding information processing in the brain.

1.7.1 Neural representations

One of the most fundamental concepts within the field of neuroscience is the idea that information is encoded by patterns of activity within neuronal populations. The precise information encoded by a particular neuronal population can range from simple visual properties such as contrast and orientation within V1 (Hubel, 1963), to planned actions within supplementary and premotor cortex (Wolpert and Ghahramani, 2000), to

highly abstracted information about expected future properties of the world and their violations (e.g. prediction errors – Friston, 2010). The type of information encoded by each of these neuronal populations is termed a neural representation.

The neural code within any given neuronal population is defined as the mechanism by which that information is coded. The rate of neuronal firing is widely viewed as the principle means of information coding in the brain, although temporal processes are likely to play an important role as well (Shadlen and Newsome, 1994; Singer and Gray, 1995; deCharms and Zador, 2000). It is important to note that there is much debate over exactly what kinds of representation are actually present within the physical substrates of the brain, as well as debates over the precise neural code of those representations. However, despite these disagreements, the fundamental view that information is encoded by some form of activity within distinct neuronal populations is not under debate, and indeed is the central tenet underlying neuroscience as a scientific endeavour.

1.7.2 Neural computations

Another key concept is that of a neural computation, which describes the transformation of information as it passes from one neuronal population to another. For example, the simple act of reaching for a cup of coffee requires a highly complex set of computations, starting from the visual system extracting information about location, shape, and size of the cup, to the high-level action planning systems which transform this information into a general goal (reach, grasp, pick up), to the next level of action planning

which plans the specific sequence of muscle movements required to achieve this goal, and finally the primary motor regions involved in directly triggering this sequence of actions. Each of these processing levels receives the input from the previous level, and performs some form of computation on this information before passing the transformed information to the next level. Neural computations are usually conceptualised as occurring through the interaction between neuronal populations within a particular processing level (e.g. through lateral inhibition between V1 neurons – Shapley et al., 2007), although back-projections from higher processing regions are widely acknowledged to have a strong influence on information processing, and the influences of these top-down connections can also be explicitly modelled as a part of any given computation (Friston, 2010).

1.7.3 Neural processes

A neural process is closely related to the concept of a neural computation. The most general definition of a neural process is as a particular type of neural activity that leads to a change in behaviour or mental state. A neural process can therefore also be thought of as a transformation of neural information – for example, object recognition is a neural process, and can be thought of as a transformation between the pattern of stimulation in the retinal cells, and the neural state underlying final recognition (although note that exactly what this neural state entails is contentious – e.g. Montaldi and Mayes, 2011; Wixted and Squire, 2011). Another way of putting it is that a neural process is the neural correlate of a cognitive process, which describes a cognitive transformation between one mental state and another.

The difference between a process and a computation is in the details – the term neural computation generally refers to a specific type of transformation that occurs within a single neural population, and that can be precisely mathematically described. Neural processes, on the other hand, generally refer to much more general neural transformations, and tend to be less precisely defined. One relevant example from the episodic memory literature is the distinction between episodic retrieval, which is a neural process involving a widely distributed network of regions (see section 1.3.2), and pattern completion, which is a specific computation that forms a critical part of episodic retrieval, occurring specifically within hippocampal CA3 (see section 1.8.2 for more details). This example shows the importance of drawing a distinction between neural processes and computations.

1.7.4 Memory traces

The fourth important definition relates directly to the theories I will describe in the next section. A memory trace describes the physical population of inter-connected neurons which is involved in the representation of a specific memory upon retrieval. Thus, while a memory representation describes the information encoded by the *activity* of a particular neuronal population, a memory trace refers to the neuronal population itself, whether it is active or not. While a full memory trace will always involve neuronal populations that are widely distributed across the brain (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011), I will also refer to hippocampal memory traces, which describe the neuronal populations specifically within the hippocampus.

1.8 How are episodic memories represented?

I now turn to the central question of my thesis – how is episodic information represented in the human brain, and specifically the hippocampus? What is the underlying neuronal architecture of an individual episodic memory trace? In this section I first discuss the current state of knowledge regarding the neural representation of episodic memory, and the limitations of current approaches. I then turn to theoretical models of episodic representation, and assess the evidence for these models. I finish this section by discussing the critical missing link between the theory and the evidence regarding the representation of episodic memories.

1.8.1 What do we know so far?

Because episodic memory can only be studied with certainty in humans, we are limited to two main sources of information regarding the link between episodic memory and the hippocampus – lesion studies and neuroimaging. As described earlier, there is unequivocal evidence from both of these sources that the hippocampus is critical for the process of episodic encoding and retrieval (at least for memories that are only a few years old). However, for the purposes of investigating individual memory representations, both of these approaches are severely limited. It is obviously not possible to measure information about representations that are no longer present, as in the case with hippocampus amnesia, so lesion studies cannot help with this question. What about neuroimaging approaches such as fMRI? While

standard fMRI approaches have proved fruitful for investigating memory processes, they do not allow the investigation of neural activity at the level of specific representations. The reason for this is that the standard approach to fMRI analysis involves searching for global activations within gross anatomical regions (e.g. the hippocampus). If we used this approach to compare two specific episodic memories, we would expect the hippocampus to be equally active for both, and we would therefore not find any useful information. Thus, while we might reasonably infer that regions that are active during episodic recall are likely to contain specific episodic representations, we cannot directly access those representations with standard fMRI.

The only method that has thus far been able to directly investigate individual episodic representations in the human hippocampus is the recording of neuronal activity from electrodes implanted within the MTL of epilepsy patients undergoing clinical assessment. To date, only one study has done so - Gelbard-Sagiv et al (2008) recorded from the MTL while the patients viewed 46 distinct movie clips (e.g. an episode of the Simpsons, Tom Cruise giving an interview). They found that the majority of neurons recorded showed a significant response to at least one of the clips, and many showed responses to multiple clips. Most importantly, however, in a subsequent free recall session, they found that the hippocampal neuronal activity reflected the activity profile during the original movie clip viewing. In other words, a neuron that responded selectively to The Simpsons at encoding, also showed selective activation at free recall. This property was only seen in the hippocampus and entorhinal cortex, and not in any of the

other locations that were examined.

This study therefore demonstrates that individual hippocampal neurons can show selective activation both at encoding and at free recall of episodic information, thus indicating that these neurons may form part of an episodic memory trace. However, it is worth noting that, although the patients viewed less than 50 movie clips, the majority of hippocampal neurons responded to more than one clip. Given the vast amount of information that is processed by the hippocampus on a daily basis, this level of neuronal selectivity would lead to each neuron responding to many thousands of events and stimuli. This clearly indicates that information (and hence episodic memory traces) cannot be coded by the selective activation profile of individual neurons. Instead, in order to cope with a realistic magnitude of data, memory traces must be represented by populations of neurons. It will therefore be important to consider population-level activity in order to develop our understanding of episodic representations. Another important limitation of this single-unit recording approach is the fact that it involves an invasive procedure in a patient population, which is far from ideal as a standard methodological technique. Overall, therefore, empirical data regarding the representation of episodic memories is severely lacking.

1.8.2 Computational theories of episodic memory

Despite the lack of empirical data, influential theoretical models of hippocampal function have been developed over the last 40 years which provide detailed predictions regarding episodic representations and computations. These theories I will refer to under the umbrella term

"computational theories", as they share many core assumptions regarding information processing within the hippocampus.

This approach to understanding memory originated with the work of David Marr (Marr, 1971), who hypothesised about the computational functions of the hippocampus based on its unique neuronal architecture. The essence of Marr's theory was that the hippocampus can rapidly and flexibly associate disparate cortical representations into a single hippocampal memory trace, which he termed a "simple representation". This sparse memory trace is synaptically linked to each of the cortical regions involved in the full representation of the memory. The result of this architecture is that partial or incomplete cues can activate the whole hippocampal memory trace, which in term activates the entire distributed cortical memory representation, thereby automatically and rapidly retrieving the entire memory. This important process has since been called "pattern completion". It is easy to see how this particular neuronal architecture may be useful for episodic memory, which is inherently complex and multimodal, and must necessarily rely on widely distributed cortical regions for a full representation of any given event.

Based on the detailed anatomical mapping of Lorente De No and Cajal (Cajal, 1911; Lorente De No, 1933, 1934), Marr assigned specific computations to the different subfields of the hippocampus (see Figure 6 for the anatomy of the hippocampal subfields, and Figure 7 for a guide to subfield connectivity). He proposed that the simple memory trace itself was created and stored within the CA fields of the hippocampus. He further

suggested that the dense recurrent collateral connections found within CA3 makes this particular subfield ideally suited to pattern completion. Marr also suggested that the dentate gyrus (DG) is likely to be involved in the creation of sparse memory representations from information projected from the entorhinal cortex. These sparse representations are then projected into CA3, and from there, into the other CA fields to be stored as a sparse, simple memory trace.

This model has since been extended through further theoretical work, and there is now broad agreement about the core anatomical and functional contributions of the different hippocampal subfields to memory (Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Information is proposed to project from the entorhinal cortex (EC) to the DG, CA3, and CA1 via the perforant path. Of these regions, CA3 is proposed to be particularly important for the rapid formation of complex associative memories (such as episodic memory) due to its unique recurrent architecture (Treves and Rolls, 1994; Rolls, 2010). The DG is thought to be critical for a process known as "pattern separation", whereby overlapping input patterns (i.e. an event which may be very similar to a previously witnessed event) are orthogonalized into distinct, sparse representations. At the same time, the component elements of an episode are projected separately into CA3 from the EC. The combination of the auto-associative CA3 architecture, and the sparse, pattern separated input from DG leads to the formation of a conjunctive episodic memory trace within CA3.

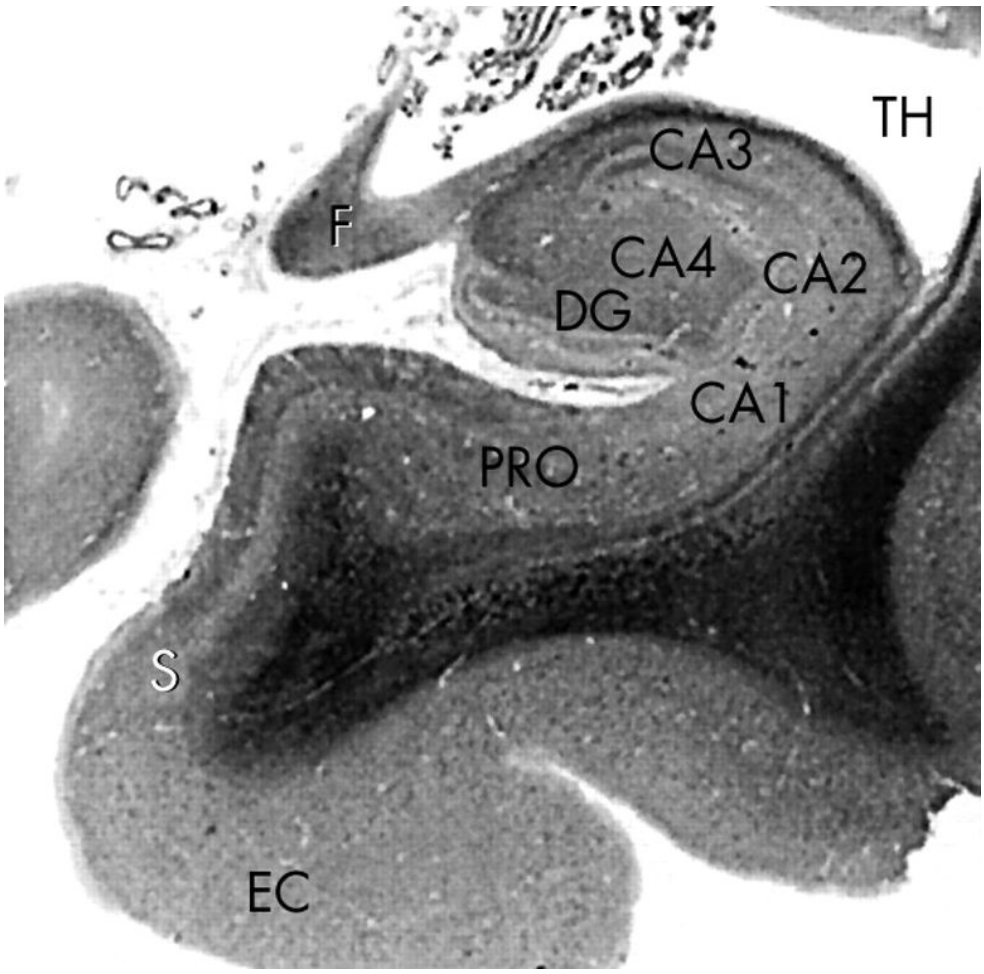

Figure 6. The subfields of the hippocampus. *This figure displays a coronal section from the right medial temporal lobe of a human brain. EC = entorhinal cortex, S = subiculum, PRO = prosubiculum, DG – dentage Gyrus, F =fimbria, TH = temporal horn of the lateral ventricle. Taken from Marshall et al. (2004) with permission from BMJ Group.*

From there, CA3 projects into CA1 via the Schaffer collaterals. The function of CA1 is less well-characterised, but there are two major theories. According to one view (Treves and Rolls, 1994; Rolls, 2010), because CA3 contains information about both the individual episodic elements as well as their associations, CA1 memory traces are required for the formation of stable, conjunctive episodic representations. The second theory suggests that CA1 is necessary for the formation of a sparse, invertible mapping between CA3 and the EC, and that this is necessary in order to avoid interference

effects between different memory traces (McClelland et al., 1995; O'Reilly et al., 2011). These two views are clearly not incompatible, and may simply reflect different perspectives on essentially the same set of computations. The critical point is that both theories suggest that CA1 contains conjunctive episodic memory traces, and that these CA1 representations are crucial for accurate retrieval of an episode.

This model so far describes how an episodic memory can be rapidly and automatically stored as an episodic memory trace within the hippocampus, through the complex interactions between the hippocampal subfields. This arrangement is also ideally suited for rapid retrieval of memories from partial cues, which will automatically trigger the original CA3 representation through pattern completion. This will, in turn, drive the CA1 memory trace, resulting in a cascade of activation that reinstantiates the entire distributed memory representation throughout the cortex. Figure 7 displays the complex pattern of connections between the subfields, and between the hippocampus and neocortex.

Notably, the concept of consolidation forms a key component of the computational models, although the theories tend to be somewhat ambivalent about how consolidation is expected to affect episodic memory (Rolls, 2010; O'Reilly et al., 2011). Indeed, the Complementary Learning Systems model makes a very similar argument to the Multiple Trace Theory, proposing that true episodic memories never become fully independent of the hippocampus (O'Reilly et al., 2011). Ultimately these models depend on empirical data to inform the consolidation debate, the confused state of

which I have already discussed in section 1.4. I therefore devote the rest of this section to a discussion of the evidence for the more general aspects of the computational theories such as pattern completion and pattern separation, setting aside the issue of consolidation.

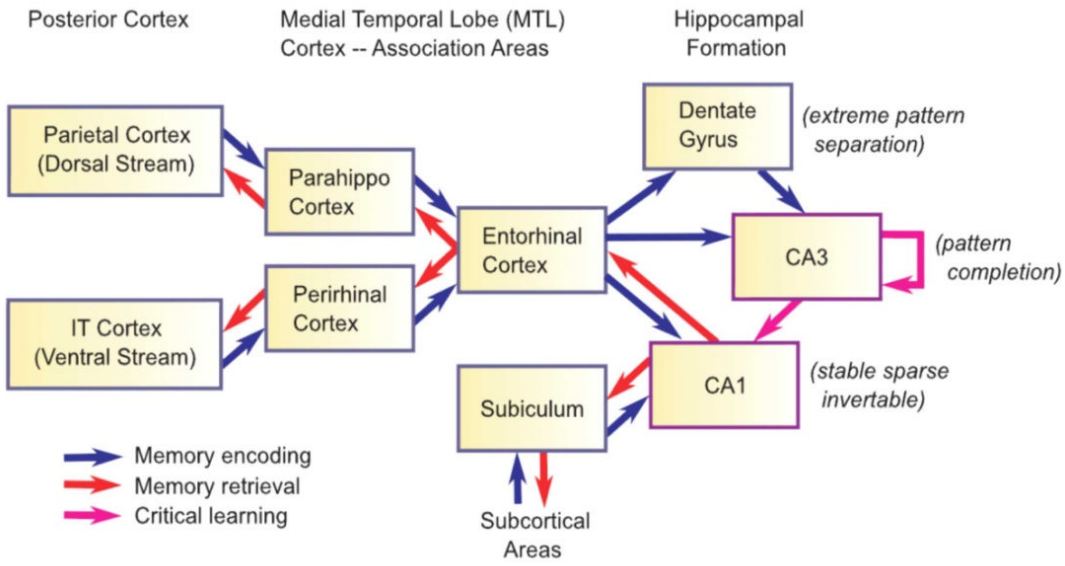

Figure 7. Connectivity of the hippocampal subfields. *This diagram shows the pattern of connections between the hippocampal subfields, parcellating the sections of the diagram into those that are proposed to be most important during pattern separation and encoding (blue arrows), and those that are critical for retrieval and pattern completion (red arrows). The interaction between CA3 and CA1 is proposed to be important for the formation of stable memory traces, here termed "critical learning" (pink arrows). Also depicted are the connections from the hippocampus to neocortex via the entorhinal cortex. From O'Reilly et al. (2011) with permission from Wiley.*

1.8.3 Empirical support for the computational theories

In recent years some experimental evidence has started to emerge for these kinds of neuronal computations within the hippocampal subfields of rats, and here I will outline some of the most relevant studies. Vazdarjanova and Guzowski (2004) used a genetic imaging method to assess the pattern of activity expressed during the exploration of various environments in regions

CA1 and CA3 of rats. By using two different methods of labelling neurons, they were able to measure neural activity that occurred (a) 2-15 minutes and (b) 25-40 minutes prior to sacrificing the animal. They placed each animal in two different environments which shared certain features at these two different time-points, which then allowed them to investigate the pattern of neural activity expressed for each environment separately. They used this elegant design to directly assess the level of overlap between the population of neurons that were active in each of the two environments within both CA1 and CA3, in order to test the assumptions of the computational theories outlined above. Across a variety of different environmental manipulations, they found that the overlap in activity was greater in CA3 than CA1. In other words, the neural representation changed less in response to small environmental changes, which is consistent with the proposed role of CA3 in pattern completion – because the second environment is similar to the first, CA3 pattern completion leads to the expression of a very similar pattern of activity for both environments.

A second study published just one day after Vazdarjanova and Guzowski (2004) found strikingly similar results when recording activation of neuronal populations in CA1 and CA3. Lee et al. (2004) allowed rats to learn two circular enclosures which had a combination of distal and proximal cues. They implanted tetrodes into CA1 and CA3, and then interleaved the previously learned “standard” enclosure set-ups with mis-match trials in which the enclosure cues were rotated to various different degrees. They found that population representation of the environment was more stable and coherent within CA3 than CA1. In other words, the patterns

of neuronal firing between the standard and mis-match environments were more highly correlated in CA3 than in CA1. This provides a second convincing source of evidence of pattern completion processes occurring within hippocampal subfield CA3.

A third study from the same year provided further evidence of a functional dissociation between CA1 and CA3. Leutgeb et al. (2004) implanted tetrodes in CA1 and CA3 of rats in order to record the pattern of activity when the animals were exposed to enclosures with varying geometric similarity (large square, small square, small circle). They repeated this in three different rooms (A, B, and C), and found that the CA3 activity was distinct across all three rooms regardless of the similarity of the enclosure. CA1, on the other hand, showed a significant degree of overlap between rooms, and this overlap increased depending on the similarity of the enclosure. These results suggest that CA3 is capable of maintaining distinct representations across environments which share some features (in this case, geometric features), while CA1 shows a more graded responses, whereby the similarity in activation depends on the degree of environmental similarity. The authors argued that these results are fully consistent with a role of CA3 in pattern separation, as proposed by the computational theories.

The evidence does indeed lead to this conclusion, but how can we account for the fact these results show the opposite effect to the previous two studies (Lee et al., 2004; Vazdarjanova and Guzowski, 2004)? All of these studies use similar paradigms, where small changes were made to the environments, and patterns of activity were measured in both CA1 and CA3. So why do

the former two studies find pattern completion in CA3, while Leutgeb et al. (2004) find pattern separation? The authors of all three studies point out that CA3 has the potential to pattern separate or complete incoming information, and there may be constant tension between these two processes. Exactly which process is ultimately deployed will depend on a variety of factors, such as the current goal-state of the animal, and the degree of overlap between the incoming information and stored CA3 representations. While there was no explicit goal-state required in any of these paradigms (animals were always freely foraging), they did use different environmental manipulations, so it is entirely possible that the environmental changes in the Leutgeb et al. (2004) study were larger than in the other two studies, thus evoking pattern separation processes within CA3. While this is a plausible explanation, what is sorely lacking here is any well-defined concept of exactly what kinds of changes should be considered more "similar" than others. What kinds of environmental or stimulus changes are likely to induce pattern separation versus pattern completion? Even more poorly understood is the relationship between goal-states and these processes. The motivation of an animal is likely to interact with the representations within the subfields in a significant way, and it will be important to gain a better understanding of these interactions.

Despite these outstanding issues, this set of studies clearly demonstrates that regions CA1 and CA3 are dissociable in terms of their patterns of neuronal activity, and implicate CA3 in both pattern separation and pattern completion. As such, the results are consistent with the computational models, and provide the first strong evidence that these models have a

genuine neurobiological grounding. The dentate gyrus (DG) is also proposed to be important for pattern separation, and a study by Leutgeb et al. (2007) provided the first experimental evidence supporting this claim. They recorded from DG and CA3 neurons while rats were exposed to a variety of "morph" enclosures which covered a range of intermediate shapes between a square and a circle. The pattern of activity in DG proved to be extremely sensitive to even small changes in the environment, showing very little overlap in the pattern of activation between similar environments. Region CA3 proved to be more robust to small changes, but showed little overlap over larger changes. Overall, the activity in both regions was consistent with pattern separation, with CA3 also displaying pattern completion properties over smaller environmental changes.

Together, these studies provide compelling evidence that the neural computations originally proposed by Marr (1971) may indeed be taking place within the hippocampal subfields of the rat during exploration of environments. But a critical question is whether there is any evidence that similar processes may also be taking place within the human hippocampus? While a few fMRI studies have reported results consistent with hippocampal pattern separation (Kumaran and Maguire, 2006; Bonnici et al., 2012), the first evidence of functional specialisation within the subfields of the human subfields came from a study by Bakker et al. (2008). They used high-resolution fMRI in combination with a Blood-oxygenation level dependent (BOLD – see Chapter 2 for physiological basis of BOLD) adaptation paradigm in order to probe the response profiles of the hippocampal subfields in human participants. BOLD adaptation is a well-known

phenomenon whereby the functional response to a given stimulus is significantly reduced on repetition. It has been proposed that this effect reflects an underlying neural adaptation effect, whereby the repeated stimulation of the same neuronal population will lead to reduced responses (Grill-Spector et al., 2001, 2006; Kourtzi and Kanwisher, 2001). Bakker et al. (2008) used this property to indirectly assess the relative impact of pattern separation/completion in each hippocampal subfield.

In order to do this, they presented a series of object pictures to participants during scanning. Some of these trials were slightly different pictures of previously seen objects (lures), and some were direct repeats of previously seen pictures. Thus, on each trial a picture could be either (a) novel (b) an exact repetition, or (c) a lure. The authors reasoned that, if a given subfield is engaged in pattern separation, then a lure stimulus should be treated like an entirely novel stimulus, despite the high degree of overlap with the original picture. If, however, the region is engaged in pattern completion, the lure is more likely to be treated like an exact repetition. Although all regions displayed activity profiles that were a mixture of completion and separation, they found that DG/CA3 showed a bias towards separation, while CA1 showed a bias towards completion. A second study by Lacy et al. (2011) replicated these results using a very similar paradigm. They also extended the results, and demonstrated that the differential pattern separation effect between DG/CA3 and CA1 is only found for highly similar lures, and disappears when the lures are more distinct (Lacy et al., 2011). These results suggest that human DG and CA3 are particularly important for pattern separation processes, consistent with both theoretical models (Treves

and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011) and animal data (Leutgeb et al., 2004, 2007).

However, note that Kumaran and Maguire (2009) have argued that there may be other possible explanations for the results of these two studies. For instance, the neural representations of objects in CA1 might be more conceptual and abstracted compared to CA3/DG, leading this region to treat lures and exact repetitions equivalently. Alternatively, the two subregions may not differ in terms of computations, but may differ in the types of information represented, such that only CA3/DG represents the relevant configural changes between original stimulus and lure. Thus, while the evidence is suggestive that pattern separation processes occur within human CA3/DG, we cannot rule out these alternative explanations entirely based on these datasets.

1.8.4 What does this tell us about episodic memory?

This set of studies has produced converging evidence that both the rodent and human hippocampus may be performing computations such as pattern separation and completion. These results therefore provide encouraging empirical support for computational accounts of memory function. However, one of the explicit goals of computational models is to provide a mechanistic explanation for the hippocampus' critical role in episodic memory (Rolls, 2010). Thus far the evidence for the theoretical models comes from either spatial tasks in rodents (Lee et al., 2004; Leutgeb et al., 2004, 2007; Vazdarjanova and Guzowski, 2004) or implicit object discrimination tasks in humans (Bakker et al., 2008; Lacy et al., 2011), and

both of these are a long way from true episodic memories. There is, therefore, a critical empirical gap between episodic memory and the computational accounts of hippocampal function. As discussed in section 1.8.1, current methods for investigating episodic memory do not allow the investigation of information at the level of episodic representations, and it is this limitation that has severely hampered attempts to provide evidence of a link between the computational accounts and episodic memory.

So where does that leave us? It is clear that, despite the strong theoretical background detailed above, we actually have very little concrete knowledge about the neural representation of episodic memories. The central aim of this thesis is to use novel analysis tools to directly investigate the representation of episodic information within the human hippocampus. By doing so, I hope to provide some new insights regarding the biological basis of episodic memory, as well as addressing some of the important debates within the field. In the next section I will argue that novel multi-voxel methods for analysing fMRI data might prove sensitive enough to allow the investigation of episodic memories at the level of individual memory traces, which would allow me to address these important issues.

1.9 Multi-voxel pattern analysis

For twenty years, scientists have been attempting to localise the neural correlates of a wide range of cognitive functions within the human brain using fMRI. Throughout this time, the mass-univariate method has dominated data analysis. This approach involves creating a model of the

experimental design that is fitted to the fMRI BOLD response at each voxel independently (a voxel is the smallest unit we can measure in a 3D brain image volume), the aim being to find activity in individual voxels that consistently shows a relationship with the experimental design. Mass-univariate analysis has served fMRI well. Nevertheless, over the last number of years there has been increasing interest in alternative methods that exploit the intrinsically multivariate nature of fMRI data. The motivation for this change stems from the belief that there may be information present in the distributed pattern of activation across voxels that is missed when looking at each voxel independently as in the mass-univariate method (Haynes and Rees, 2006; Norman et al., 2006). This type of multivariate approach is commonly known as multi-voxel pattern analysis (MVPA), or decoding (Figure 8). I will use both of these terms interchangeably throughout this thesis (although see Kriegeskorte (2011) for a technical discussion of the distinction between encoding and decoding MVPA approaches).

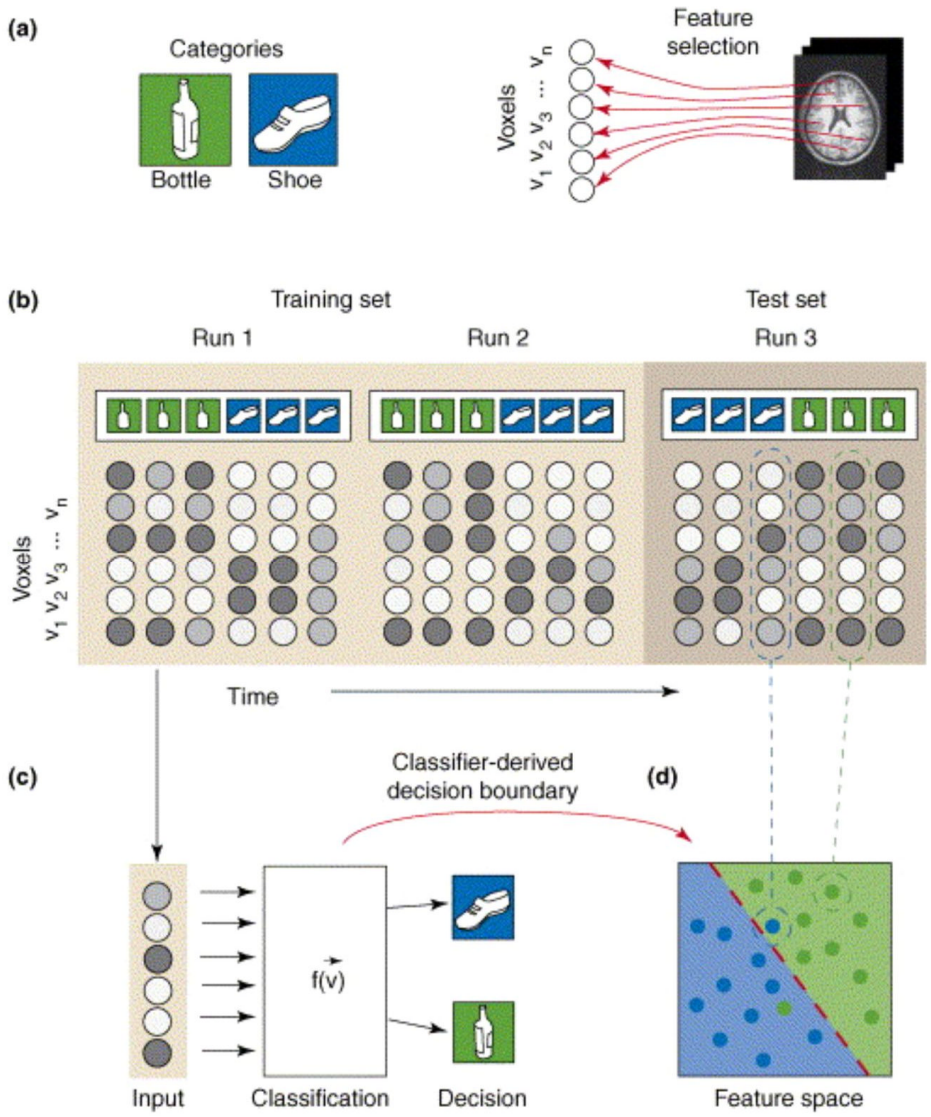

Figure 8. The principles of multi-voxel pattern analysis. *(a) In this example, a participant views stimuli from two objects categories, bottles and shoes. A 'Feature selection' procedure is used to select a set of voxels which will be included in the classification analysis. For now, let us assume that this simply involves creating an anatomical region of interest within lateral occipital cortex (LOC), and we include all voxels from within that ROI (for more on feature selection, see Chapter 2 - Methods). (b) A summary of fMRI activation for the presentation of each trial is created. From this, we can look at the pattern of activation across LOC voxels for each "bottle" trial and each "shoe" trial. This is depicted for six example voxels, for nine trials of each category. These are divided into a training set and a test set. (c) In order to assess the information that may be present within these patterns of activation, a multi-voxel classifier can be trained to differentiate the two categories based on all of the trials from the training set. (d) The classifier will extract statistical regularities in the multi-voxel pattern of activation for each category, and use this to optimize a decision boundary (red dotted line) that best separates the two categories within the high-dimensional space of the voxel patterns. Each dot corresponds to a single data trial, and the colour indicates its category. The background colour of the feature space indicates the predicted category of all the trials on that side of the*

decision boundary. To test the classifier, the trained classifier is presented with new data from the independent test set. Each of these trials is projected into the feature space, and is classified as a shoe or a bottle depending on which side of the decision boundary it falls on. In this example we can see that both trials have been classified as being bottle trials. Overall, if the number of trials which are successfully predicted is significantly greater than chance, then we conclude that there must be a significant degree of information present within the pattern of activation across multiple voxels.From Norman et al. (2006) with permission from Elsevier.

A clear demonstration of the potential of MVPA was provided by Haxby et al. (2001), who found that neural representations of object categories, such as places and faces, were more widely distributed and overlapping within the ventral temporal cortex than had been thought previously. Importantly, they examined specific regions where the individual voxels (using a mass-univariate approach) responded strongly to one category or another, and found that within these supposedly category-selective regions, there still existed considerable information in the distributed pattern of activation about the non-preferred categories. This illustrates the complementary nature of the information offered by mass-univariate and MVPA analyses, and suggests that MVPA may be more sensitive to the presence of information about specific representations such as object categories. Since this early study, MVPA has been applied in a wide range of cognitive domains including perception (Haynes and Rees, 2005; Kamitani and Tong, 2005), emotion (Peelen et al., 2010; Baucom et al., 2012), decision-making (Kahnt et al., 2011), and memory retrieval (Polyn et al., 2005).

Importantly, MVPA has been shown to be sensitive to information about highly specific representations, such as the orientation of gratings (Haynes and Rees, 2005; Kamitani and Tong, 2005) based on V1 activation. This

level of sensitivity suggests that the technique is able to detect information based on subtly different patterns of underlying activity at the level of neuronal populations within a region. MVPA may, therefore, prove to be sensitive enough to detect information at the level of individual episodic memory representations. However, is it possible to detect such subtle representations from noisy patterns of activity from within the hippocampus?

A study by Hassabis et al. (2009) used an MVPA approach in order to investigate the representation of spatial information within the human hippocampus. As discussed in section 1.5, the hippocampus has long been known to play a crucial role in the representation of space, and particularly allocentric spatial location as exemplified by the existence of “place cells” in both rodents (O’Keefe and Dostrovsky, 1971; O’Keefe and Nadel, 1978) and humans (Ekstrom et al., 2003). Hassabis and colleagues therefore aimed to determine whether an MVPA approach would be sensitive enough to this kind of subtle neuronal information to allow the decoding of four specific locations within a virtual environment.

Participants controlled their movement within a virtual room while undergoing scanning, and were required to navigate between all four corners of the room in a pseudo-random order. This was repeated across multiple blocks within two separate virtual rooms. Importantly, whenever the participants reached a target location, there was a period where their view within the virtual room automatically tilted down to look at a patch of carpet which was visually matched across the four target locations; there was then a visual countdown to the start of the next trial. The activity from these

periods was extracted and used in the MVPA analysis, meaning that the direct visual input was exactly matched for each of the four locations within each room. Nevertheless, they found that it was possible to decode these four locations from patterns of fMRI activity across voxels in the hippocampus in each of four participants. These results demonstrate that highly abstracted representations of space are present and detectable from patterns of fMRI activation within the human hippocampus. Importantly for my purposes, this study also demonstrated that the hippocampus is a viable target for MVPA studies.

Put together, these previous studies suggest the possibility of decoding episodic memories from patterns of activity within the human hippocampus. Such an approach would give us privileged access to information about individual episodic memory traces, and would provide us with an exciting opportunity to examine important properties of episodic memory. This therefore provides a clear motivation for the application of MVPA to human episodic memory, which is the major theme of my thesis. In the next section I will give an overview of my specific aims, and the experiments I have conducted in pursuit of those aims.

1.10 Thesis overview

The overarching aim of this thesis is to develop a more comprehensive understanding of the neural basis of episodic memory through the investigation of episodic representations in the human hippocampus, and the processes underlying the formation and retrieval of those representations.

In the first four experiments, I used a combination of high-resolution fMRI and MVPA which allowed me to investigate the representation of episodic memories at the level of individual memory representations. While I have already discussed the basic concept behind MVPA and the rationale for its use in the investigation of episodic memory (see section 1.9), I include a more in-depth discussion of the advantages and disadvantages of MVPA along with some methodological and conceptual issues to be aware of in Chapter 2. In this chapter I also describe the basics of fMRI, the BOLD signal, and univariate analysis.

In Experiment 1 (Chapter 3) I used MVPA to investigate episodic representations within the hippocampus and surrounding MTL. No previous study had yet provided evidence that such complex information could be decoded using fMRI, so the major focus of this study was to determine whether or not this was possible. In Experiment 2 (Chapter 4), in a joint study with Heidi Bonnici, we took this approach further and used MVPA to decode autobiographical memories from the recent past and from over a decade ago, in the remote past. The central aim of this experiment was to investigate the strength of episodic memory representation within the hippocampus and neocortex as a function of time in order to test some key assumptions of the Standard Consolidation Theory (Squire, 1992; Squire et al., 2004). If the hippocampus showed no decline in the strength of memory representation between the recent and remote memories, then this would be strong evidence against the view that consolidation leads to hippocampally-independent memory traces.

Experiment 3 (Chapter 5) was designed to further probe the nature of the episodic memory trace within the hippocampus. Various accounts of the hippocampus propose that one possible reason for its crucial role in episodic memory is its ability to create unique, conjunctive representations from overlapping inputs (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Eichenbaum, 2004). In order to directly test this proposal, I asked participants to recall a set of overlapping movie clips while in the fMRI scanner. I then used MVPA analysis to determine whether the hippocampus contains unique memory traces of such highly overlapping episodes. In Experiment 4 (Chapter 6), I extended my examination of this dataset in order to investigate the representation of episodic memory traces within the individual subfields of the hippocampus. The computational theories of episodic memory (Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011) make further specific hypotheses regarding the contribution of each individual hippocampal subfield to episodic memory, and this experimental design provided an excellent opportunity to directly test these theories. I therefore collected sub-millimetre high-resolution structural MRI data from the same set of participants, and used this to manually segment the hippocampal subfields for each participant. Using these new regions of interest, I conducted a further MVPA analysis in order to investigate the nature of the episodic memory traces within each specific subfield. In so doing, I hoped to provide the first empirical link between theoretical computational processes within the hippocampus, and complex episodic memory.

Episodic representations in the hippocampus may depend on a process known as scene construction (Hassabis and Maguire, 2007, 2009; Hassabis et al., 2007a). In my final study I used a slightly different approach in order to further explore the role of the hippocampus in the construction of scenes, so that we might begin to understand the mechanisms by which this process contributes to the formation and retrieval of episodic memories. A recent study demonstrated that amnesic patients show deficits in a cognitive phenomenon known as boundary extension (Mullally et al., 2012), which is thought to depend on the automatic, implicit construction of scenes beyond the borders of a given view (Intraub, 2012). This study therefore suggests that hippocampal damage not only impairs the ability to actively and explicitly construct mental scenes, but also impairs the automatic and implicit construction of extended scenes that ordinarily occurs whenever we perceive a scene. In Experiment 5 (Chapter 7), I used a standard fMRI paradigm in order to investigate the neural correlates of boundary extension, thereby allowing us to investigate the role of the hippocampus in automatic scene construction. By better defining the scene construction processes taking place within the hippocampus in this way, I hoped to aid the development of a more mechanistic understanding of scene construction. This level of explanation will be a crucial step if we wish to truly understand how scene construction contributes to the representation of episodic memories.

In Chapter 8, I provide a general discussion of the results from this set of experiments. I focus on the implications of each study for our understanding of episodic memory and the role of the hippocampus, with a particular

emphasis on how these results impact current theories and debates within the literature. Finally, I include a brief discussion of what I consider to be the critical outstanding questions in the field, along with some proposed directions for future research.

1.11 Publications

The following publications have arisen from work described in this thesis:

Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA. 2010. Decoding individual episodic memory traces in the human hippocampus. *Current Biology* 20:544–547.

Chadwick MJ, Hassabis D, Maguire EA. 2011. Decoding overlapping memories in the medial temporal lobes using high-resolution fMRI. *Learning & Memory* 18:742–746.

Chadwick MJ, Bonnici HM, Maguire EA. 2012. Decoding information in the human hippocampus: A user's guide. *Neuropsychologia* (in press).

Chadwick MJ, Mullally S, Maguire EA. The hippocampus extrapolates beyond the view in scenes: an fMRI study of boundary extension (under review).

Chadwick MJ, Bonnici HM, Maguire EA. CA3 size predicts individual differences in the perceived distinctiveness of overlapping episodes (in preparation).

Bonnici HM, **Chadwick MJ**, Hassabis D, Lutti A, Weiskopf N, Maguire EA. Decoding recent and remote autobiographical memories informs systems-level consolidation (under review).

Bonnici HB, **Chadwick MJ**, Kumaran D, Hassabis D, Weiskopf N, Maguire EA. Multi-voxel pattern analysis in human hippocampal subfields (under review).

Bonnici HM, Sidhu M, **Chadwick MJ**, Duncan JS, Maguire EA. Memory representations in temporal lobe epilepsy: an fMRI multi-voxel pattern analysis study (under review).

Work also undertaken during my PhD but not reported here:

Bonnici HM, Kumaran D, **Chadwick MJ**, Weiskopf N, Hassabis D, Maguire EA. 2012. Decoding representations of scenes in the medial temporal lobes. *Hippocampus* 22(5): 1143-1153.

Bonnici HM, **Chadwick MJ**, Maguire EA. Autobiographical memory representations in human hippocampal subfields (in preparation).

Bonnici HM, **Chadwick MJ**, Maguire EA. Cortical representations of episodic memories immediately after acquisition – an MVPA study (in preparation).

2 Chapter 2

Methods

2.1 Methods overview

The aims of this chapter are three-fold. First, I give a general overview of the types of tasks and methodologies that I applied in the five experiments reported in this thesis. Second, for methods that are shared across multiple experiments, I provide their details here, while details that are specific to each experiment are included in the relevant experimental chapter. Finally, the use of fMRI in general, and MVPA decoding analysis in particular, form a critical part of the thesis. As such, I present an account of the biophysics underlying MRI imaging, and the general concepts involved in mass-univariate and MVPA approaches to fMRI data analysis. These latter sections comprise a comprehensive discussion of the various methodological and conceptual issues surrounding the application of MVPA to fMRI. These issues are critical for ensuring that MVPA is used appropriately, and I have given each of them careful consideration during the analysis of my experiments.

2.2 Participants

All participants were healthy and right-handed. They were recruited through the WTCN or UCL Psychology Department experiment recruitment pool. Further details of each specific group of participants are provided in the methods section of the relevant experimental chapter.

2.3 Experimental Tasks

A range of tasks were used in the experiments reported in this thesis, and were specifically designed to capture the naturalistic aspects of episodic memory. This involved eliciting vivid autobiographical memories from both recent and remote time-points, creating video clips of naturalistic “episodes”, employing green-screen technology to create overlapping episodic movie clips, and using photographs of everyday scenes. Full details of each task are included in the methods section of the relevant experimental chapter.

2.4 The biophysics of MRI

All of the studies presented in this thesis used functional magnetic resonance imaging (fMRI) to indirectly measure neural activity whilst human participants performed cognitive tasks. In the next sections I will outline the basic principles of MRI in general, and fMRI in particular. I will then provide the details of the specific MRI scanners and sequences used in my experiments. Following this I will describe the physiological mechanisms proposed to underlie fMRI measurements, otherwise known as the BOLD response, and discuss implications for the kinds of inference we can and cannot make based on BOLD activity.

2.4.1 MR signal generation

Under normal background conditions, thermal energy causes the single proton in a hydrogen nucleus to spin about itself (Jezzard et al., 2003). Because hydrogen carries a positive charge, this spin generates a magnetic

field, referred to as the “magnetic moment” of that nucleus. In the absence of a strong magnetic field, the spins of hydrogen protons are randomly oriented, and over any meaningful population of hydrogen atoms, the resulting magnetic fields will tend to cancel each other out. However, if a magnetic field is applied externally, the protons will align their orientation to that field, initiating a gyroscopic motion known as precession.

Precessing protons can be in two states, based on the orientation to the magnetic field – parallel or anti-parallel (see Figure 9), which is partially determined by the orientation of spin at the onset of the external magnetic field. The parallel state requires less energy than the anti-parallel state, and is therefore more stable. Consequently, there will be more protons in the parallel than anti-parallel state in the present of a magnetic field, with the relative proportion of the two states dependent on the temperature and strength of the magnetic field. In my experiments, all MRI data was collected using a 3 Tesla static magnetic field.

When a proton in the anti-parallel, high-energy state, falls into the parallel, low energy state, a photon is released which contains an amount of energy equal to the difference between the two states. Conversely, a proton with parallel spin can jump to the higher-energy state by absorbing a photon with the required amount of energy to bridge the difference between the two states. For a given type of atomic nucleus and magnetic field strength, it is possible to calculate the frequency of electromagnetic radiation required to change protons from a low to high-energy state. This is known as the Larmor frequency, and these properties form the core of MRI technology.

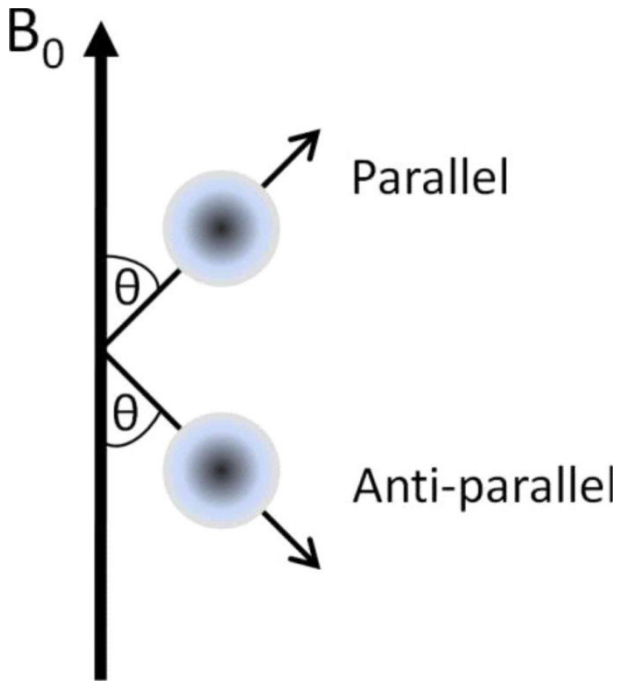

Figure 9. Magnetic spin. *Precessing hydrogen nuclei can be in either a parallel (low-energy) or anti-parallel (high-energy) states relative to the static magnetic field (B0).*

In an MRI scanner, radiofrequency (RF) coils bombard protons with energy in the form of photons. Some of this energy is absorbed by the protons, which causes those protons to jump to a high energy, anti-parallel state. Over time, the photons are then released again by those same protons as they revert to the more stable, low energy state. The signal from the released photons can be detected by a receiver RF coil as it gradually decays over time. This decay period is known as spin relaxation, and generally occurs within a few seconds. There are two types of spin relaxation – longitudinal and transverse. For a given substance (e.g. water or fat) in a magnetic field of a given strength, the rates of these two types of spin relaxation are given as time constants. Longitudinal relaxation is the type just described above, whereby the protons revert from a high to a low-energy state, and the time-

constant associated with longitudinal relaxation is called T1. The RF excitation pulse also causes a degree of coherence between spins precessing around the main field vector, as they begin their precession within the transverse plane at the same starting point. Over time, however, this coherence begins to decay, and the spins become out of phase with one another. This signal decay is known as transverse relaxation, and the time constant associated with this is called T2. Importantly, in addition to this intrinsic T2 decay, field inhomogeneities can also lead to a loss of spin coherence, and the combined effects of both of these causes leads to a signal loss known as T2* decay, characterised by the time constant T2*.

2.4.2 MR image formation

Based on these signal properties and time constants, we want to be able to acquire meaningful three-dimensional images of brain structure or function. In order to achieve this, a second type of magnetic field, called the “gradient field” is applied. Overall, three magnetic fields are applied along three axes. First of all, a static field is applied along the z axis, which selects a single plane from which to collect data (the “slice select” gradient). Two further gradient fields are applied to this slice, which allow distinct spatial locations to be encoded by both the frequency and phase of the detected signal from that slice. The signal frequency is determined by the frequency encoding gradient field applied in the x axis. This is then followed by the phase encoding field applied along the y axis. This process is repeated for all the slices within the selected brain volume, and together this dataset allows the separation of signal into three dimensional volume elements or “voxels”. Each voxel represents the distinct signal received from that location in space,

and the resolution of the voxels depends on the particular MRI sequences used (typically ranging from 1.5mm to 4mm resolution).

2.4.3 MR scan types

There are two important parameters that determine the type of MRI image collected. The first is repetition time or TR, which is the interval between two consecutive 90° RF pulses. The second is echo time or TE, which is the time between the initial RF excitation and data acquisition. At short TR and TE intervals, the T1 characteristics of tissue are emphasised, which results in a T1-weighted image. Conversely, longer TR and TE intervals will instead emphasise the T2 relaxation time, leading to T2-weighted images. Both of these types of image can be used for studying the structure of the brain and, as I will explain in the next section, T2*-weighted images can be used to study brain function. In all experiments described in this thesis, standard T1-weighted structural images were acquired, while in Experiment 4 I also collected high-resolution T2-weighted structural images.

A key challenge in the acquisition of fMRI data is the fact that whole-volume images must be acquired rapidly, in around 3 seconds. In order to achieve this, I used Echo-planar imaging (EPI) sequences (Mansfield, 1977; Weiskopf et al., 2006), which allow the collection of an entire slice by changing spatial gradients rapidly following an RF pulse.

2.4.4 The BOLD signal

Red blood cells contain a type of protein called haemoglobin, which transports oxygen in the blood. This allows the oxygen to be transported around the body, and transferred into tissue wherever oxygen is required. If haemoglobin is bound to oxygen it has no magnetic properties, whereas unbound haemoglobin is paramagnetic. That means that deoxyhaemoglobin alters a magnetic field into which it is introduced. Together, this means that the magnetic state of blood reflects its level of oxygenation. Critically, therefore, paramagnetic molecules such as deoxyhaemoglobin induce local magnetic field inhomogeneities. As described earlier, these inhomogeneities influence the T2* decay time. This means that the ratio of oxyhaemoglobin to deoxyhaemoglobin directly effects the T2* parameter, which provides an indirect measure of the metabolic state at each voxel. The type of signal measured using this approach is called the blood-oxygenation-level dependent (BOLD) contrast, which is the difference in signal on T2*-weighted images as a function of the amount of deoxyhaemoglobin. Work in animal models and in humans demonstrated that this BOLD contrast can be reliably detected using MR methods (Ogawa and Lee, 1990; Ogawa et al., 1990, 1992).

The relationship between the BOLD response and underlying neural activity can be characterised by the haemodynamic response function, or HRF (Figure 10), which consists of three main phases. First, the increase in neural activity within the given region leads to an increase in metabolic demand, which rapidly leads to an “initial dip” in the BOLD response, as oxygen is used up faster than it can be replaced (Menon et al., 1995; Duong

et al., 2001). Second, the increased metabolic demand then leads to an increase in the supply of oxygenated blood to the region through its local vasculature, as the capillaries dilate. This leads to a large increase in BOLD signal, peaking around six seconds after the onset of activity. Third, once the BOLD peak has subsided, there is an "undershoot" that lasts for several seconds. FMRI analysis relies upon identifying the clear peak in HRF response, as the initial dip is smaller and more difficult to identify (Heeger and Ress, 2002).

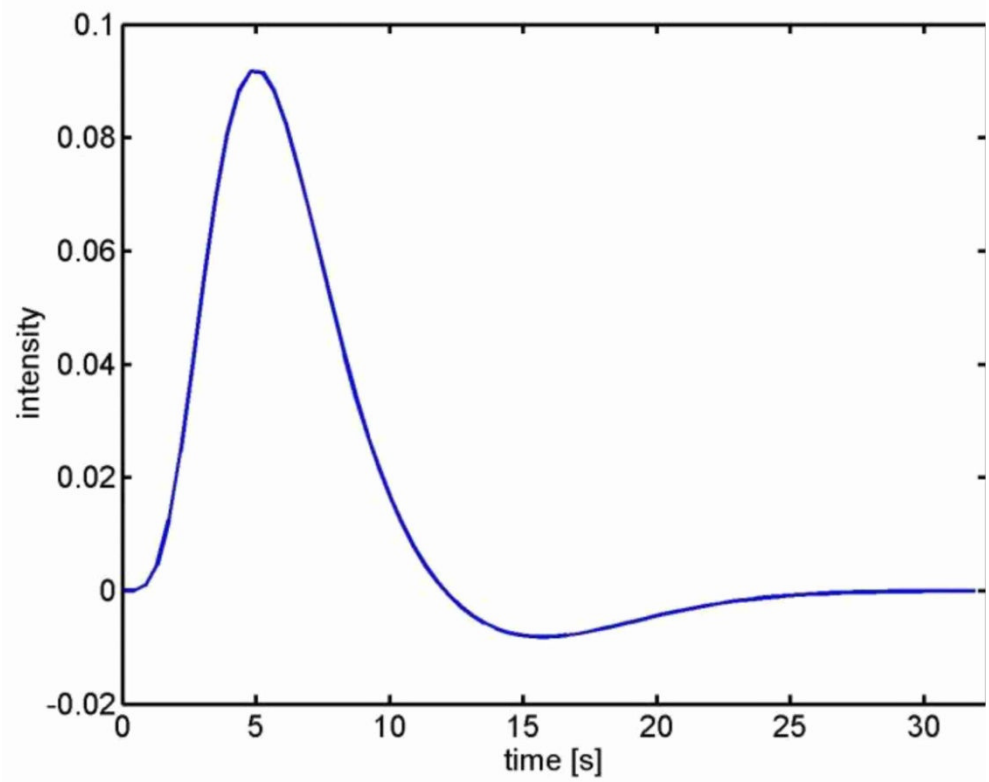

Figure 10. Canonical haemodynamic response function. *x-axis: Time (seconds). y-axis: Amplitude of response (arbitrary units). Response peaks at around 6 seconds after initial stimulation.*

So far I have discussed the HRF and BOLD response in terms of underlying "activity", but what kind of neural activity is actually reflected by the BOLD signal? Work conducted by Nikos Logothetis over many years has

provided evidence that the BOLD signal does not directly reflect neuronal spiking activity (although of note it usually correlates quite highly with spiking), but instead reflects oscillatory activity, particularly within the gamma band (Goense and Logothetis, 2008; Logothetis, 2008; Magri et al., 2012). As noted briefly in Chapter 1 (section 1.7.1), the representation of information within the brain is currently thought to depend on a combination of both neuronal spiking and temporal coding, of which gamma-band oscillations are one of the most important proposed mechanisms (Shadlen and Newsome, 1994; Singer and Gray, 1995; deCharms and Zador, 2000). Thus, the measurement of gamma activity may be just as important for inferring the underlying representation as the measurement of neuronal spiking. This work therefore validates the use of fMRI for measuring meaningful neuronal activation, and allows us to make inferences about neuronal representations from the BOLD signal.

2.4.5 Resolution of fMRI

While the dynamics of the underlying neuronal activity occur on the time-scale of milliseconds, the BOLD response takes a number of seconds to evolve, peaking at around six seconds. Due to this lack of temporal resolution, fMRI is not used to investigate complex temporal dynamics in the brain. Instead, it is used to assess the spatial location of activation that occurs in response to a given stimulus, regardless of exactly when that neural response occurs. The spatial voxel resolution of fMRI is determined by the specific MRI parameters chosen. However, note that the actual spatial resolution achieved is generally coarser than the voxel size, due to the fact that the BOLD response critically depends on the local vasculature.

2.5 Specific MRI details

2.5.1 MRI scanners

All functional imaging data that I describe in this thesis were acquired on a 3T Magnetom Allegra head-only MRI scanner (Siemens Medical Solutions) operated with a standard transmit-receive head coil. The sub-millimetre structural data in Experiment 4 (Chapter 6) were acquired on a 3T Magnetom TIM Trio whole body MRI scanner (Siemens Medical Solutions) operated with the standard transmit body coil and 32-channel receive head coil.

2.5.2 MRI sequences

2.5.2.1 fMRI

Two fMRI sequences were used during this thesis. For the four MVPA decoding studies, a high-resolution ($1.5mm^3$) sequence was used for two main reasons. First, this spatial resolution is high enough to permit reasonably precise apportioning of functional signal to two anatomically neighbouring regions. This was particularly important for Experiment 4 (Chapter 6), where I was interested in extracting signal from the specific hippocampal subfields. The second reason is that higher resolution allows one to sample more voxels within a given region, which potentially provides greater sensitivity to detect underlying multivariate information (see section 1.7.8 for more details). However, note that the use of this high-resolution sequence entailed focussing data acquisition on a partial volume through the MTL, in order to maintain a reasonable TR. In Experiment 5, I was

interested in investigating univariate activity across the whole brain, and therefore used a standard resolution sequence instead ($3mm^3$).

The specific high-resolution sequence was a T2*-weighted single-shot echo-planar imaging sequence (in-plane resolution = 1.5 × 1.5 mm^2, matrix = 128 × 128, field of view = 192 × 192 mm^2, 35 slices acquired in interleaved order, slice thickness = 1.5 mm with no gap between slices, echo time [TE] = 30 ms, asymmetric echo shifted forward by 26 phase-encoding lines, echo spacing = 560 μs, repetition time [TR] = 3.5 s, flip angle α = 90°). All data were acquired at 0° angle in the anterior-posterior axis in one single uninterrupted functional scanning session. An isotropic voxel size of 1.5 × 1.5 × 1.5 mm^3 was chosen for an optimal trade-off between BOLD sensitivity and spatial resolution. Furthermore, the isotropic voxel dimension reduced resampling artifacts when applying motion correction.

The standard resolution sequence was a T2*-weighted single-shot echo-planar imaging sequence which has been optimized to minimize signal dropout in the medial temporal lobe (Weiskopf et al., 2006). The sequence uses a descending slice acquisition order with a slice thickness of 2mm, an interslice gap of 1mm, and an in-plane resolution of 3 x 3 mm. 48 slices were collected in order to cover the entire brain, resulting in a repetition time of 2.88s. The echo time was 30 ms and the flip angle 90°. All data were acquired at a -45° angle to the anterior-posterior axis.

2.5.2.2 Fieldmaps

For all experiments, fieldmaps were acquired with a standard manufacturer's double-echo gradient echo field map sequence (TE = 10.0 and 12.46 ms, TR = 1020 ms, matrix size = 64 × 64), with 64 slices covering the whole head (voxel size = 3 × 3 × 3 mm^3). These were used in subsequent distortion correction, or "unwarping" (Hutton et al., 2002).

2.5.2.3 Structural images

For all experiments, T1-weighted high-resolution 3D modified driven equilibrium Fourier transform (MDEFT) whole-brain structural MRI scans were acquired for all participant after the main scanning session with 1 mm isotropic resolution (Deichmann et al., 2004). Additionally, Experiment 4 required the acquisition of high-resolution, sub-millimetre structural images for the purposes of hippocampal segmentation. These were acquired in a separate session in a 3T Trio MRI scanner (see scanner details above), in a partial volume focused on the temporal lobes. A single-slab 3D T2-weighted turbo spin echo sequence with variable flip angles (Mugler et al., 2000) combined with parallel imaging was employed to simultaneously achieve a high image resolution of ~500 μm, high sampling efficiency and short scan time while maintaining a sufficient signal-to-noise ratio (SNR). After excitation of a single axial slab the image was read out with the following parameters: resolution = 0.52 x 0.52 x 0.5 mm^3, matrix = 384 x 328, partitions = 104, partition thickness = 0.5 mm, partition oversampling = 15.4%, field of view = 200 x 171 mm^2, TE = 353 ms, TR = 3200 ms, GRAPPA x 2 in phase-encoding (PE) direction, bandwidth = 434 Hz/pixel, echo spacing = 4.98 ms, turbo factor in PE direction = 177, echo train

duration = 881, averages = 1.9. For reduction of signal bias due to, e.g. spatial variation in coil sensitivity profiles, the images were normalized using a pre-scan and a weak intensity filter was applied as implemented by the scanner's manufacturer. To improve the SNR of the anatomical image, four scans were acquired for each participant, co-registered and averaged.

2.6 Univariate analysis of fMRI data

As outlined above, fMRI represents a powerful tool for investigating neural processes in the human brain in vivo. Here I describe the various analysis steps required to make these kinds of inferences. Although the main analysis method implemented within this thesis is a multivariate decoding approach, I will first describe the standard mass-univariate approach to fMRI analysis. It is important to understand this for several reasons. First, various pre-processing step can be applied for both types of analysis. Second, both methods face some of the same methodological issues, and it is informative to consider how these issues are addressed by each approach. Thirdly, in order to fully understand why multivariate analyses can in some cases be advantageous, or even necessary, it is first important to understand the limitations of the univariate approach. I therefore provide in this section a description of this type of analysis.

2.6.1 Analysis overview

Here I will give a brief description of the various pre-processing and analytical steps carried out in a standard mass-univariate analysis. This particular analysis pipeline is the default pipeline for Statistical Parametric

Mapping (SPM, specifically, SPM8), which was used throughout this thesis for all standard imaging analyses. This general pipeline is common to most software packages that use the voxel-wise parametric mapping approach. Following this outline, I will provide a more detailed description of each of the processing steps in turn.

The first step required is to discard the initial volumes acquired in each scanning session, in order to allow for T1-equilibration effects (Frackowiak et al., 2004). This is the case for all studies described in the thesis, and I always removed the first six volumes. The next set of steps is required in order to move, or "warp" the functional data from all subjects into alignment with each other. These steps I will describe in the "spatial preprocessing" section, but the aim here is to make sure that the data for each subject are in approximate anatomical alignment with one another. This is necessary to assign an observed response to a particular brain structure or cortical area at the group level.

Following this preprocessing, statistical analysis is applied to each individual participant individually. This analysis is based on fitting a general linear model (GLM) to each voxel independently (Frackowiak et al., 2004). The output of this analysis is then taken to the second level, using a summary statistic approach, whereby the first-level analysis for each participant is summarised in a single contrast image. Classical statistics are then applied to each voxel at the second level, and a critical part of the voxel-wise approach involves the appropriate family-wise error correction for multiple comparisons (i.e. the fact that the statistical test has been

applied to many thousands of voxels across the brain).

2.6.2 Spatial preprocessing

2.6.2.1 Realignment

The functional images for each participant are first realigned into a common reference frame in order to correct for any head movements during scanning, and to ensure that all images are anatomically aligned. Realignment involves the estimation of six parameters (three translation and three rotation parameters) which between them describe the movement of an image within three dimensions, and the rotation in three dimensions. This procedure then minimises the differences between the scans in order to optimise anatomical alignment.

2.6.2.2 Unwarping

Another potential source of noise is the contribution of magnetic field inhomogeneities, and the interaction between movement and these inhomogeneities. In order to correct for these sources of noise, it is possible to use a procedure known as “unwarping”. This involves the collection of separate images which map out the field inhomogeneities (known as fieldmaps), and using them to generate a forward model of movement-by-inhomogeneity interactions (Andersson et al., 2001; Hutton et al., 2002). In all experiments in this thesis, realignment and unwarping steps were applied to the functional data as part of the preprocessing stream.

2.6.2.3 Spatial normalisation

After within-subject spatial processing, it is necessary to transform the images from all participants into a standard stereotactic space, in order to map functional activations to specific anatomical locations at the group level. This is achieved by geometrically distorting (warping) each subject's brain into a standard shape. This can be achieved by directly warping each subject's functional data to a standard functional template that is within Montreal Neurological Institute (MNI) space. Alternatively, it is possible to use T1-weighted structural images to guide normalisation. This is achieved by first co-registering the structural image to the mean of the (unwarped, realigned) functional volumes. This step uses a 6-parameter realignment algorithm in order to "move" the structural image into alignment with the functional data. SPM's segmentation protocol is then applied to the structural image, which uses an iterative approach to segmenting the structural data into different tissue types (usually grey and white matter and CSF), and simultaneously uses the tissue types to optimise the structural normalisation into standard MNI space. The resulting warp information from this procedure can then be applied directly to the functional data in order to normalise it to MNI space. A potentially more powerful approach is offered by a newer normalisation algorithm in SPM called DARTEL. This involves first optimising the warp between the structural images within your dataset. The result of this can then be transformed into MNI space. Overall this approach appears to give more accurate normalisation results (Ashburner, 2007), and this is the method I used for normalisation in my standard whole brain Experiment 5.

In the four MVPA decoding experiments, all analyses were performed on regions of interest within the native space of each subject's functional data. Therefore no spatial normalisation was required (the exception to this is the "information map" analyses reported in Experiments 1 and 2 – see individual experimental chapters for more details).

2.6.2.4 Smoothing

The final preprocessing step in a standard SPM analysis involves the application of a Gaussian smoothing kernel to each voxel of the functional data. The extent of the smoothing is determined by the full-width at half-maximum (FWHM) of the kernel, and is typically between 6 and 12mm in fMRI studies. Smoothing has several important effects on the data. First, despite the normalisation process applied to the data of each subject, there will still be small differences in the anatomical location of each voxel across subjects. Secondly, the statistical inference applied at the group-level, or second-level, analysis within SPM depends critically on Gaussian random field theory. Random field theory requires the underlying data to be smooth for valid inference, meaning that smoothing is an essential step within SPM (Frackowiak et al., 2004). Notably, as I will discuss in greater detail later on, this step may be removing important information that is present in the unsmoothed patterns of data across voxels, and this is one potential advantage of the MVPA approach.

2.6.3 Mass-univariate statistical analysis

In the mass-univariate approach, the time-series of each voxel is analysed independently. Within each voxel, a General Linear Model (GLM - see below) is fitted to the data, and the resulting statistics can be displayed as a statistical parametric map (SPM). Using the SPM, it is then possible to apply classical statistics to determine whether there are any regionally specific effects related to the experimental variable modelled within the GLM.

2.6.3.1 The General Linear Model (GLM)

The GLM is an equation that expresses the observed response variable Y in terms of a linear combination of explanatory variables X (which are entered in the form of a design matrix), plus an error term (Frackowiak et al., 2004):

$$Y = X\beta + \varepsilon$$

Within this equation, β is a vector containing the parameters that are to be estimated. In the analysis of fMRI data, the observed response variable Y is the time series of observed BOLD signal at the given voxel. Note that the GLM approach assumes that the residuals are independently and identically distributed, which is not the case for fMRI time series. The HRF is of longer duration than the typical scan acquisition time, which leads to serial correlations among error terms. SPM accounts for these autocorrelations by imposing a known temporal smoothing function on the time-series and adjusting the estimators and degrees of freedom accordingly (Frackowiak et al., 2004).

The design matrix X is the user-specified component of this analysis, which consists of columns which are referred to as regressors. The set of regressors applied within a GLM will represent the experimental manipulations, confounds and covariates of no interest. Each of these regressors specify the onset of trials (or other events to be modelled) for each experimental session separately. Each trial or event can either be modelled as occurring with a certain duration (specified by the user), which is modelled with a square wave or "boxcar" function, or as a transient event, which is modelled as a stick (delta) function. In order to produce a model of the proposed BOLD response, each of these regressors is convolved with the chosen HRF function to take into account the temporal profile of the BOLD response. The standard choice of function is the canonical HRF function (Frackowiak et al., 2004). Each convolved regressor will therefore capture the proposed BOLD profile associated with a given explanatory variable. Note that typically, regressors that track the movement of a subject's head during scanning are included as regressors of no interest, in order to remove any confounding effects. This kind of regressor is not convolved with the HRF response.

2.6.3.2 Model estimation and statistical inference

Once the GLM is specified, it must be estimated at every voxel within the functional volume. This process estimates the best linear combination of regressors in order to provide an optimal fit with the time series at that voxel. The standard approach uses a maximum likelihood estimate, which results in the generation of parameter estimates (also known as betas) for each regressor, for each voxel. The beta for a particular explanatory variable will

therefore represent the extent to which the time series at that voxel fits that explanatory variable. In other words, if a voxel tends to show no change in BOLD in response to a particular stimulus, then the beta for a regressor relating to that stimulus will be very low, and vice versa.

Statistical inference can then be made based on both the betas themselves and their estimated variances. It is possible to run two types of test at this point – an F-test, or a t-test. The former tests the null hypothesis that all specified betas are zero, and produces an F-statistic. The latter tests the more specific null hypothesis that some particular linear combination (e.g. a subtraction) of betas is zero, and produces a t-statistic. The linear combination of betas is user-specified by a set of contrast weights – for example, a specification of [1 -1] would be used to look at a differential response evoked by the first two regressors in the design matrix. The t-statistic can then be calculated by dividing the contrast of betas by the standard error of that contrast. This error term is estimated using the variance of the residuals about the least squares fit. After this procedure is applied to all voxels in the functional volume, the "SPM" is created, based on the full set of t or F statistics (Frackowiak et al., 2004).

2.6.3.3 Multiple comparisons

This approach involves many thousands of statistical tests across the whole brain volume, which leads to a severe multiple comparisons problem. For example, a typical fMRI experiment may include 20,000 voxels across the brain, and if we were to use a standard statistical threshold of $p<0.05$, we would expect 5% of those voxels to pass this threshold just by chance. This

would lead to 1000 “active” voxels even if there were actually no underlying effect. We therefore need to have some level of control against the likelihood of finding false positives among so many data points.

One simple way to do this is to use a Bonferroni correction, which divides the desired statistical threshold α (e.g. $p<0.05$) by the total number of independent tests being conducted. However, in extreme cases such as fMRI, where there are tens of thousands of statistical tests, this tends to create an unfeasibly high statistical threshold – i.e. it is highly conservative. Notably, however, this form of correction is only appropriate when all of the statistical tests are genuinely independent of one another. This turns out not to be the case in fMRI, as the betas will tend to be highly correlated across spatially adjacent voxels. This is due to the fact that activity for any given task will tend to cluster within certain anatomical locations, and this clustering is enhanced by applying the spatial smoothing during preprocessing.

For this type of data, Bonferroni correction is not appropriate, and instead, an alternative approach based on Random Field Theory (Frackowiak et al., 2004) can be applied. Essentially, Random Field Theory makes explicit use of the fact that fMRI data will be spatially clustered, and corrects the statistical threshold based on the number of spatial clusters (resolution elements, or resels), rather than individual voxels. This approach will therefore lead to lower corrected values of α than Bonferroni correction, while still providing a good level of control for false positives.

The specific implementation within SPM allows multiple comparisons to be

applied in two different ways, each of which provides a different sort of inference about the underlying neuronal activation. First, a “height” threshold can be applied in which the multiple comparisons correction is based on the “height” of the statistical “peaks” of each spatial cluster. This approach allows for inference based on highly localised activation at a particular point (or points) within the brain. The second approach is to use “cluster level inference”, which is based on the size of spatial clusters that survive a certain threshold. This approach requires an initial statistical threshold (typically $p<0.001$, although this is arbitrary, and any threshold could potentially be used), and the inference is then based on determining whether any of the clusters surviving this threshold are larger then would be expected by chance (again, after correcting for multiple comparisons).

The above procedures are appropriate for more exploratory analyses, when no specific anatomical regions are hypothesised a priori. If a particular experiment is strongly hypothesis driven, then it is possible to use a more restricted multiple comparisons correction known as small volume correction (SVC). In order to do this, a specific anatomical region or regions must be specified, and SPM will apply Random Field Theory correction within this reduced functional volume (Frackowiak et al., 2004). I applied this approach to the whole brain standard fMRI analysis described in Chapter 7 in order to investigate activity in regions that were specified a priori.

2.6.3.4 Group level analyses

The analysis procedures described above all refer to the single subject level, and serve to provide betas and contrast estimates for each individual subject. How then do we make inferences at the group level? In fact, many of the principles already described also apply to the group (or second) level analysis, due to the "summary statistic" approach used. This approach is implemented in a two-stage procedure, where the first stage involves generating contrast estimates for each subject as described above. The set of contrast "images" from the group of subjects is then treated as a new response variable Y in the second-level GLM analysis. The design matrix at the second level can be used in the same way as the first, and various different types of regressor can be added here (although the general rule is to keep this level as simple as possible). Thus, the simplest possible second-level analysis (and a very common one) is a one-way t-test, which tests the null hypothesis that the contrasts across all subjects are zero. Once estimated, the second-level analysis will generate an SPM, and the statistical analysis of this SPM is conducted in exactly the same way as described above, including the mandatory correction for multiple comparisons. This general approach is one example of a random effects analysis, which allows inferences to be made about the general population from which the sample of subjects was drawn. This stands in contrast to fixed effects analyses, which assume homogeneity across subjects, and can therefore be more biased by outlier effects.

2.6.3.5 fMRI experimental design

There are two basic types of experimental design used within fMRI studies – "blocked" and "event-related". In a blocked design, subjects engage in particular type of task repeatedly for an extended period time (e.g. around 30 seconds or so). Blocks of different types of task are alternated, or randomised, and the BOLD responses to each type of block can be compared at analysis. Blocked designs are highly efficient in terms of the power to detect changes in the underlying BOLD response (Frackowiak et al., 2004). However, in many cases it is more desirable to investigate neural activity on a trial-by-trial basis, in which case event-related designs are used. In this thesis, the multivariate decoding experiments all used a slow event-related design, where each trial lasted around 25 seconds in total, with the vivid recall period of interest typically lasting between 6 and 12s. The final experiment, however, used a more standard event-related design in order to differentiate BOLD responses on a trial-by-trial basis.

2.7 Multi-voxel pattern analysis

As outlined in my introductory chapter, the main focus of this thesis is on the investigation of episodic information at the level of individual memory traces, as to date very little is known about the nature of these representations. While classical mass-univariate approaches to fMRI analysis are insensitive to this level of information (as I will explain in more detail below), various recent studies have demonstrated that multi-voxel pattern analysis (MVPA) is sensitive enough to detect subtle differences in underlying neuronal representations (Haynes and Rees, 2005; Kamitani and

Tong, 2005; Hassabis et al., 2009). There are two main reasons for the increased sensitivity of MVPA. Firstly, by taking a multivariate approach to analysis, MVPA can combine information that is weakly present in many different voxels in order to detect a robust signal that is not present within any individual voxel by itself (Haynes and Rees, 2005, 2006; Kamitani and Tong, 2005). Secondly, as discussed earlier, standard univariate processing involves spatial smoothing of the data. This is an important step for standard analysis, as it will boost the signal associated with the spatially segregated neural processes that are commonly investigated using mass-univariate analysis (see section 1.7.3 for a definition of neural process). However, in the case of subtle information such as distinct episodic representations, this process may be removing critical, fine-grained information that is present in the distributed pattern of activity across local sets of voxels.

A recent MVPA study by Hassabis et al. (2009) demonstrated that it is possible to differentiate specific spatial locations within a virtual environment based solely on the patterns of activation across voxels within the hippocampus. These results clearly demonstrate that it is possible to decode subtle, specific neural representations within the hippocampus using this approach. This suggests that MVPA may be sensitive enough to detect other types of hippocampal representations such as individual episodic memories. The primary aim of my first experiment (Chapter 3) was to determine whether it is indeed possible to decode individual episodic memories from activity within the hippocampus, and the subsequent three experiments all follow on directly from the results of this study. Given the clear importance of this method to my research, I dedicate the next section

to a more in-depth discussion of MVPA methodology.

2.7.1 MVPA methods

There are a host of different approaches to analysing fMRI data with MVPA, and to describe all of these methods in detail would require an entire thesis in itself. Instead, I will give an overview of the steps involved in a typical MVPA classification analysis, as most of these principles apply to MVPA analysis no matter which specific method is chosen. I will then provide further detail regarding the various methodological choices to be made when designing any MVPA study, along with a discussion of important conceptual issues surrounding MVPA implementation and interpretation.

2.7.2 MVPA classification overview

Figure 11 provides a schematic demonstrating the basic principles of a typical MVPA classification analysis. In this example, a subject is asked to retrieve two specific episodic memories repeatedly while undergoing fMRI scanning (panel A). Step 1 of any MVPA analysis always involves the initial selection of data (i.e. the set of voxels that serve as the input), and in this case we anatomically define a region of interest within the hippocampus (panel B). This allows us to specifically investigate episodic information contained within the hippocampus itself. Next, from each voxel within this ROI, we extract the functional data from the recall period of each trial (there are various ways of doing this, which I will describe in more detail later on), and label each trial as Memory A or B (panel C). In a typical MVPA analysis, the dataset will be divided into a training set and a test set. The classifier

algorithm (of which there are many – see later section) is first "trained" on the training dataset, during which the classifier will use some statistical learning rule to optimise the division of the two types of data trial (in this case Memories A and B). In the example given here, a linear classifier (e.g. a linear discriminant function) has been trained to discriminate the two types of memory, by determining an optimal "decision boundary" that best separates the two memories within the high-dimensional space of the voxel patterns (feature space).

Once trained, a classifier can then be "tested" on independent test data, which simply involves determining which side of the decision boundary a new trial falls on. In this example the trial falls within the green "Memory B" side of the boundary, which leads to the (correct) prediction that the memory recalled on that trial was memory B (panel D), whereas if it had fallen on the blue side, the prediction would have been incorrect. Overall, the performance of the classifier is assessed by calculating the proportion of trials from within the test dataset that are correctly classified, given as a percentage (panel E). The reason for employing this train-test approach is to circumvent over-fitting, to which complex multivariate algorithms are prone, leading them to produce inflated estimates of the underlying information. By testing the algorithms on independent test data, unbiased estimates of the model fit can be generated (Duda et al., 2001).

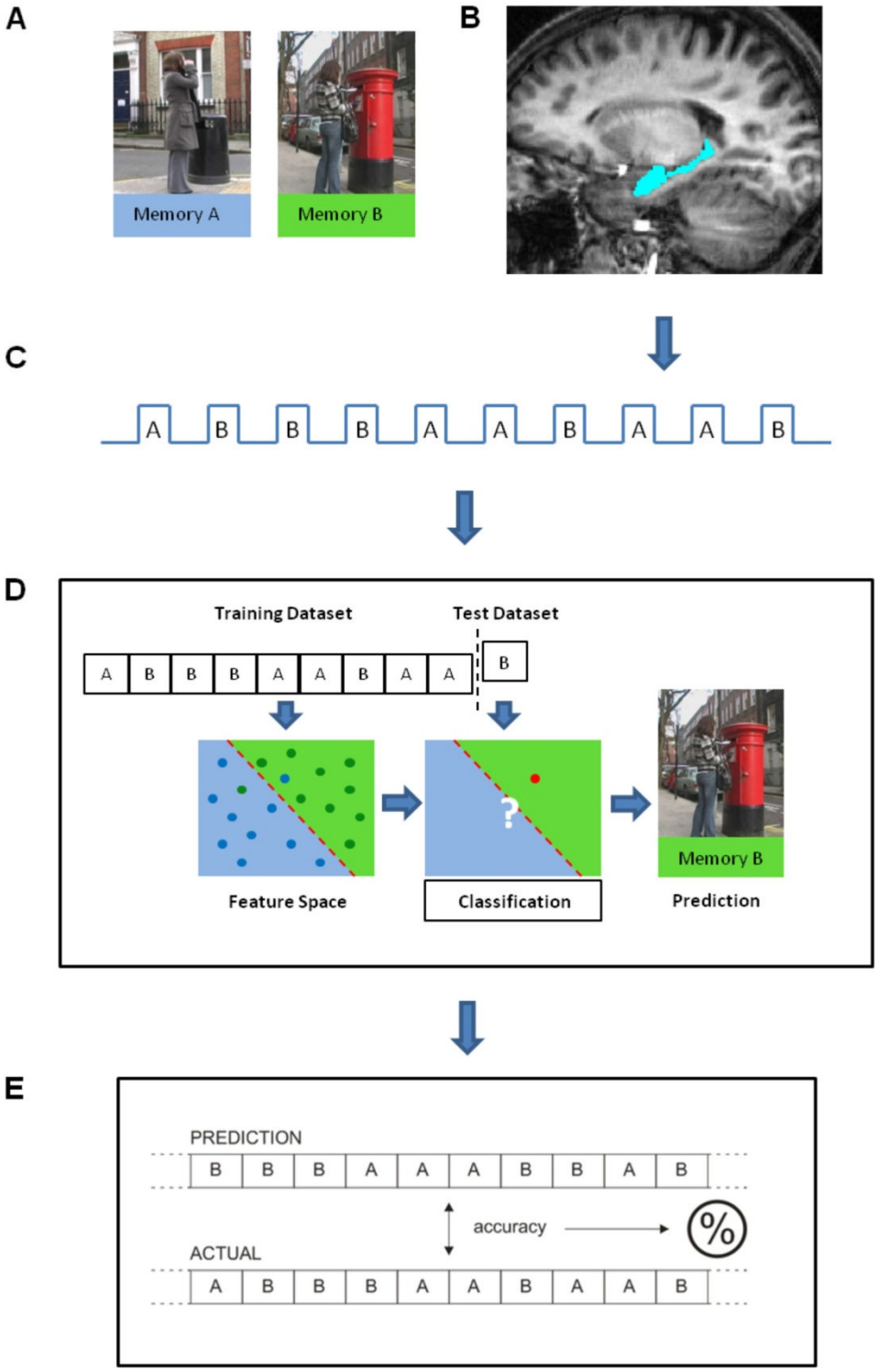

Figure 11. MVPA classification example*. (A) Subjects recall two specific memories multiple times while undergoing functional scanning. These are labelled as memories A and B. (A) A region of interest is defined in the hippocampus. (C) Functional activation is extracted for each retrieval trial, and labelled as memory A or B. (D) An MVPA classifier is trained on nine of the trials, which constitute the "training dataset", and tested on the remaining "test dataset" trial. This leads to the correct prediction that the memory being retrieved on that trial was memory B. (E) Classifier performance is determined by calculating the percentage of trials that are correctly classified.*

This example gives an idea of how a typical MVPA analysis would be conducted, but it is important to note that the field of MVPA is ever-evolving and improving as new, more refined methods are developed. The specific set of methods used throughout the thesis reflects this development – as new methods became available, I applied them to subsequent analyses. Thus, I have used Support Vector Machine (SVM) classifiers and a more sophisticated Bayesian model-based decoder (Multivariate Bayes – see Friston et al., 2008), using various types of pre-processing, and various types of feature selection. Rather than detail each of these methods here, I will provide the specific methods within each individual experimental chapter. Instead, I will use the rest of this section to discuss the various choices that must be faced when designing and analysing an MVPA study, along with some of the key conceptual issues involved in correct analysis and interpretation of the data.

2.7.3 Initial selection of voxels

One of the critical choices to be made in an MVPA analysis concerns the initial selection of data. The most common approach is to employ regions-of-interest (ROIs), where the multivariate information is assessed within a specific brain region, which can be defined either anatomically or functionally (e.g. using a functional localiser). Once the region is designated, the activity is extracted from all voxels within that ROI, and an MVPA analysis is applied in order to interrogate the patterns of information present within this set of voxels. It is important to note that ROI analyses, when based on functionally defined ROIs, require researchers to exercise care to ensure that they do not fall foul of “double-dipping”. This is where the

same data are used for selection and further in-depth analyses in a biased fashion (for further details on this error, see Kriegeskorte et al., 2009; Pereira et al., 2009). Provided this is not the case, ROI-based MVPA analyses can be extremely useful for hypothesis-driven research questions, as it allows one to draw specific conclusions about the informational content of a particular brain area, and in some cases to compare the informational content across different regions. All four of the decoding experiments presented in this thesis have used an anatomical ROI-based approach focussing on episodic representations in the hippocampus and surrounding MTL cortex.

A second approach involves investigating multivariate information that may be widely distributed across the whole brain. In these analyses the voxel activity is extracted from the whole brain (or a large part of it, such as the grey matter), and MVPA is applied to this entire set of voxels. This type of analysis is therefore not anatomically specific, but instead examines information that might be widely anatomically distributed. There are two potential problems with the use and interpretation of whole-brain decoding. First, because there are so many voxels included in the analysis, this approach will only be successful when the brain states are quite distinct, and will likely not work as effectively for more subtle information present in localised regions of the brain. This is due to the fact that analysing data at the whole-brain level vastly increases the number of data features (voxels). This leads to a substantially greater ratio of features to data points (e.g. functional volumes), which in turn leads to a greater chance of overfitting while training the algorithm (Guyon and Elisseeff, 2003; Pereira et al.,

2009). Second, determining the location of information from a whole-brain MVPA is not straightforward, due to the inherently multivariate nature of the analysis (Friston et al., 1995). If, however, the scientific question specifically concerns information about cognitive states which may be widely distributed across the brain, then whole-brain analyses may be preferred to other methods. For example, Polyn et al. (2005) used this method in a study where they found that information about stimulus category was present in the brain-wide pattern of voxel activation during free recall of individual stimulus items. Furthermore, this category-level information was actually present prior to the retrieval of individual items, indicating that some form of context-dependent retrieval may have helped the participants to recall the specific items. Importantly, the purpose of this study was to map brain-wide activity patterns to cognitive states in order to inform theories of cognition rather than to localise information within the brain. It is precisely this kind of question that can be most usefully informed by the application of whole-brain MVPA.

A third common method is known as “searchlight” analysis (Kriegeskorte et al., 2006), which is a means of assessing local multivariate information across large areas of the brain, or even the whole brain. In a searchlight analysis, a “searchlight” region of interest is created for every single voxel across the brain. Each searchlight consists of a sphere of voxels (typically around 100 voxels) surrounding the central voxel. A separate MVPA analysis is applied to each of these searchlights, creating an accuracy value for every single voxel in the brain, which can be displayed as an “accuracy” or “information map”. A statistical threshold can then be applied to this map

at either the single subject or group level in much the same way as a mass-univariate analysis. This method therefore allows one to search over the whole brain for information carried in the local multivariate response patterns. It could be considered as an intermediate between the former two methods, as it effectively applies local ROI-based MVPA analysis across the entire search space, which could include the whole brain (or the cortical surface as in recent surface-based searchlight approaches – see Oosterhof et al., 2011). This method is probably the most appropriate for exploratory analysis of representations across the whole brain (or a portion of it), and allows for the localisation of information in a way that whole-brain MVPA does not. The major draw-back of the searchlight approach is that the many thousands of MVPA analyses lead to the problem of multiple comparisons that is also inherent in mass-univariate analysis, and the same procedures should be used to correct for this.

In summary, ROI-based analyses are often appropriate for hypothesis-driven MVPA analyses or for additional MVPA analyses within the context of mass-univariate studies, but for more exploratory analyses, searchlight MVPA is preferable for its ability to accurately localise multivariate information across the whole brain. Whole-brain MVPA is most useful when the question of interest regards a mapping between cognitive states and widely distributed neural activity.

2.7.4 Selecting an MVPA method

There are many different MVPA algorithms to choose from, prompting the obvious question, which one is best? One of the most widely used

algorithms is the linear support vector machine (SVM) as it is a powerful and sensitive multivariate tool and easily accessible. This is the type of algorithm used in Experiments 1-3, employing the libsvm implementation, which is a freely available, easy to use library of SVM tools compatible with multiple platforms (Chang and Lin, 2011). Other algorithms include: linear discriminant analysis, nearest neighbour, naïve bayes, multinomial logistic regression, and classification trees, to name just a few. Several studies have directly compared the performance of different MVPA algorithms (Cox and Savoy, 2003; Mitchell et al., 2004; Ku et al., 2008; Misaki et al., 2010), the most comprehensive of which was a comparison of classification algorithms and pre-processing methods by Misaki et al. (2010). While slight advantages of some algorithms over others have been found, there is not a great deal of consistency between the datasets, suggesting that algorithm performance may depend on the particular dataset used. Overall, there is no clear evidence suggesting a strong benefit for any type of MVPA algorithm over others at this time. One notable limitation of all these techniques is the need to train the algorithm, which necessitates the repetition of the stimulus classes multiple times during the experiment. In many cases, this is not a problem, as the representations themselves are expected to remain relatively stable over multiple repetitions. However, in domains such as learning/encoding, where the dynamic change in representations over time is important, it would be desirable to investigate the representational properties of individual stimuli without having to present them multiple times. This is currently not easily accomplished using MVPA algorithms such as SVM.

Another commonly used MVPA approach is representational similarity analysis (Kriegeskorte et al., 2008a, 2008b). This is based on the simplest kind of multivariate inference one can make – taking the pattern of voxel activation elicited by two different stimuli, and measuring the multivariate distance between these two patterns using a simple measure such as a correlation. Despite the simplicity of this method, when appropriately applied this type of analysis can reveal information about the structure of representations with a good deal of flexibility. Kriegeskorte et al. (2008b) demonstrated this in a comparative study of object representations within human and monkey inferior temporal cortex. In addition to comparing different species, different techniques were used in both species, with fMRI data collected from the human participants, and electrophysiological data from the monkeys. A large number of stimuli were presented to both species, and a correlation was calculated for each pair of stimuli. This was derived from the pattern of voxels in inferior temporal cortex in humans, and from the spiking activity across multiple electrodes in the monkeys. This step effectively abstracted the information from data that was species and technique-specific, to data that was coded in terms of the similarity relationship between each pair of stimuli. This abstraction allowed them to compare the representational structure of the many stimuli (now represented as a correlation matrix, or "similarity matrix") across the species. They found a striking correspondence in the similarity matrices between the species, indicating that both humans and monkeys may code visual stimuli in a similar way within inferior temporal cortex, thus demonstrating the potential power of this approach.

Because representational similarity analysis rests on a conceptually simpler approach than the more complex MVPA SVM algorithms, it can lend itself to easier interpretation of the results, as well as potentially providing more flexibility in exploring the relationships between different representations. Another advantage is that pair-wise correlations do not require multiple stimulus repetitions. It is therefore possible to investigate the representational properties of stimuli that are presented only once, which could potentially be an advantage for certain experimental questions. On the other hand, this approach is likely to be less sensitive than more complex algorithms when investigating subtle differences in multivariate data.

One promising recent development is a Bayesian model-based approach to decoding implemented within SPM called Multivariate Bayes (MVB). MVB maps multivariate voxels responses to a psychological target variable (e.g. individual memories), using a hierarchical approach known as Parametric Empirical Bayes (Friston et al., 2008; Morcom and Friston, 2012). MVB uses the same design matrix of experimental variables used in a conventional SPM analysis. When a decoding contrast is specified, a Target variable X is derived from this contrast, after removing confounds. The multivariate voxel activity provides the predictor variable Y, which the MVB model will try to fit to X, ultimately producing a log model evidence, or Bayes factor for that model. The log evidence can be considered as a measure of the mutual information between the multivariate data and the psychological variable. There are several potential advantages to the use of this kind of model-based approach over other methods such as the SVM. Firstly, by explicitly modelling the mutual information in this way, MVB is

potentially more sensitive to the underlying neural representations. Secondly, this method does not require the train-test, cross-validation approach in order to assess the underlying information, as this is provided by the log evidence for the model. Thirdly, because the multivariate data from a region is explicitly formulated as part of the model in an MVB design, it becomes possible to directly compare information across different regions, as this now reduces to a model comparison. Thus, the MVB approach is a potentially attractive method for multivariate decoding of fMRI data.

Another powerful model-based approach to decoding has been developed over the last few years by Jack Gallant and colleagues (Kay et al., 2008; Naselaris et al., 2009, 2011; Nishimoto et al., 2011). Using this approach, they have demonstrated that it is possible to predict the appearance of novel natural scenes (Kay et al., 2008), and even YouTube movie clips (Nishimoto et al., 2011), solely on the basis of voxel patterns of activation in visual cortex. This group uses a voxel encoding approach, whereby theories about the activity of underlying neuronal populations are used to model the response of each voxel to stimuli. For example, Naselaris et al. (2009) modelled the response of voxels in early visual cortex based on evidence that early visual cortex represents visual stimulation in three low-level domains – orientation, spatial frequency, and spatial location. The tuning of each voxel to these domains was modelled using a Gabor wavelet approach. Using the fully trained model, they were able to reconstruct novel scenes solely on the basis of the pattern of activation across visual cortex. In addition to being theoretically important for testing models of neural representation, the ability to investigate the neural response to novel stimuli

provides a significant practical advantage over MVPA approaches such as SVM which, as alluded to above, must be trained on multiple repetitions of each stimulus class. However, while the voxel encoding approach has proven successful in early visual cortex where there are well-defined models of neuronal population dynamics, it will be a significant challenge to apply the same approach to higher-level regions such as the hippocampus.

Ultimately there is no right or wrong answer to the question of what MVPA method to adopt, but a pragmatic approach would be to consider how distinct the representations are likely to be – if the neural representations are expected to be relatively distinct, then a simpler multivariate approach such as representational similarity analysis may be suitable, due to its ease and flexibility of interpretation. If, however, the differences between the representations are likely to be quite subtle, then a more complex algorithm will probably be more appropriate. Given the obvious complexity of episodic memory representations, I elected to use more powerful MVPA approaches. Thus, in the first three experiments, a linear SVM was used, while in Experiment Four, a model-based decoding approach (MVB) was used in order to allow a direct comparison of the representational content of the different hippocampal subfields.

2.7.5 Data preprocessing

As well as a choice of MVPA algorithms, there are also a variety of approaches to data pre-processing. The earlier MVPA studies generally extracted the raw BOLD signal (after correcting for linear or nonlinear signal drift) at around 6s following the onset of the stimuli, and used this

raw signal as the MVPA input. While this approach has had a good deal of success, it does not attempt to model the HRF function of the BOLD response in any way, and may therefore be ignoring important information. An alternative approach is to explicitly model the HRF for each trial in a general linear model (GLM), and use the resulting parameter (beta) estimates as the MVPA input, and this approach has also proved successful in a number of studies. More recently, a comprehensive comparison of various MVPA analysis steps by Misaki et al. (2010) suggested that using t-statistics based on GLM beta estimates (by dividing the beta by its standard error estimate) produced optimal MVPA results. For a comparison of different pre-processing approaches to event-related MVPA, see Mumford et al., (2012), who also concluded that using a form of t-statistic produced optimal results. In summary, the pre-processing method of choice at present appears to be the use of the GLM to produce t-statistics as the input to MVPA analyses. However, it is important to note that this does not invalidate the use of other approaches such as raw BOLD or betas. Rather, the evidence suggests that these approaches may be sub-optimal, reducing the power of the analysis and making it more difficult to observe significant results.

The preprocessing used in each of the four decoding experiments described in this thesis reflects this gradual refinement of the field – the first two experiments used the raw BOLD signal approach, while Experiment three used the t-statistic method. In Experiment four a model-based (MVB) approach was used, which explicitly models the expected time-course of the data. Thus, no additional pre-processing was required in this latter

experiment.

2.7.6 Feature selection

Within a given dataset, it is likely that some voxels will not carry any useful information about the representations of interest, only adding noise to the MVPA analysis. A frequently used method within MVPA research is "feature selection", whose purpose is to reduce a set of voxels to those that are most likely to carry information (or inversely, to remove those voxels most likely to carry noise). There are many approaches to feature selection (Guyon and Elisseeff, 2003), and I do not intend to detail them all here. However, the majority of the available methods involve two basic steps. In step 1, the informational content of each voxel is assessed, and the set of voxels is ranked accordingly. Step 2 then involves the application of a threshold criterion to these ranked data in order to select the set of voxels most likely to contain information. Finally, this reduced set is used as the input to the MVPA analysis. In order to detect the subtle episodic information within the hippocampus, in Experiment 1 I developed a novel type of feature selection tool. This involved using a searchlight MVPA to assess the local information at each voxel within an ROI, and rank the resulting searchlight accuracies (step 1). Following this, I used a conservative threshold, and selected only the top ranked searchlight for inclusion in the final MVPA analysis (step 2). The full details of this approach are included in the methods section of Chapter 3.

A related approach to reducing noise in a dataset is "feature reduction", also known as "dimensionality reduction". In this approach, the aim is not to

remove individual voxels from the analysis, but instead to summarise the dataset in a smaller number of features using approaches such as principal component analysis or independent component analysis. In this case, the input to the MVPA algorithm is no longer the set of individual voxels, but instead the set of principal/independent components, which may help to reduce noise. However, it is important to note that neither approach to feature selection/reduction is perfect, and there is no guarantee that important information will not be lost during a feature selection or reduction step. For a thorough discussion of this issue and related methodological points, see Pereira et al. (2009).

2.7.7 MVPA versus fMRI adaptation

Another method of fMRI analysis that has been used to probe neuronal processes at the level of representations is fMRI adaptation (fMRIa). This examines the effect of repeating stimuli over time with the hypothesis that stimuli that activate overlapping neuronal representations will elicit a reduced response, for which there is substantial evidence (Grill-Spector et al., 2001, 2006; Kourtzi and Kanwisher, 2001). Recent findings suggest that MVPA and fMRIa, whilst appearing to be similar, may actually be indexing different types of underlying information. For instance, Drucker and Aguirre (2009) directly compared MVPA and fMRIa in an object shape task, and found a double dissociation within the lateral occipital complex (LOC), with ventral LOC showing adaptation effects, and lateral LOC showing decoding effects. The interpretation of these results was that decoding analyses are more sensitive to information coded by narrowly tuned neurons clustered by their response properties, whereas adaptation is more sensitive to

information coded by broadly tuned neurons with no clustering principle. Similarly, Epstein and Morgan (2012) found interesting distinctions between MVPA and fMRIa analyses of scene and landmark representations. Together, this evidence suggests that MVPA and fMRIa are not simply interchangeable approaches, and may provide complementary insights into information processing. It is worth noting that in regions involved in memory, there can often be other processes involved when stimuli are repeated, which can confound, or at least complicate the interpretation of adaptation effects. For instance, a region may show increased activity due to recognition processes upon viewing a repeated stimulus, which would have the opposite effect to those predicted by adaptation. MVPA analyses do not have this particular limitation, and thus I chose an MVPA approach for the experiments described in this thesis.

2.7.8 High-resolution fMRI and MVPA

A current debate in the MVPA literature concerns the level of information being detected by this technique (Kamitani and Sawahata, 2010; Op de Beeck, 2010a, 2010b; Swisher et al., 2010; Freeman et al., 2011), and one of the practical questions arising from this debate is whether high-resolution fMRI is necessary for MVPA analysis. There are many studies that have reported robust MVPA results using standard resolution fMRI (e.g. 3mm isotropic voxels), including some of the earlier studies (e.g. Haynes and Rees, 2005; Kamitani and Tong, 2005), which demonstrates that high-resolution is not a pre-requisite for all decoding analyses. However, the question of whether high-resolution scanning can increase the power of MVPA analyses is still an open question, and has not been fully explored

with regard to hippocampal representations.

All four decoding experiments described in this thesis used a high-resolution (1.5mm isotropic voxels) sequence focused on the MTL (Carr et al., 2010). This decision was based on the assumption that maximising the spatial resolution within the relevant region of interest will also maximise the multivariate signal of the underlying information. In Experiment 3 (Chapter 5) I will describe a control analysis where I directly tested this assumption, and provide evidence to suggest that high-resolution scanning is important for this level of representation.

2.7.9 The interpretation of classifier accuracies

There is a large amount of variability in the level of classification accuracy reported in the MVPA literature, with some studies describing impressively high classification rates of 80% or more. Thus, one common question concerns what level of accuracy should be considered meaningful. It is important to note that the level of accuracy that it is possible to achieve in any given study depends heavily on the complexity of the information being decoded. When two representations are highly separable, such as faces and places, then it should be possible to classify them with a high degree of accuracy. If, however, the representations are more complex (such as episodic memories), then the patterns of activity relating to each representation may be more difficult to separate, and the MVPA classification accuracy will consequently be lower. In some circumstances therefore, it is not reasonable to expect high levels of classification accuracy. Ultimately, however, what is relevant is not so much the absolute level of

accuracy achieved, but whether the results are robust enough to be statistically significant, and replicable.

2.7.10 Decoding different levels of information

Beyond all of the methodological choices and challenges, the most critical element of any MVPA study is proper consideration of the type of information being decoded and, from that, making appropriate inferences about the underlying neural processes and representations. In general there are three types of information that have been investigated using MVPA, the broadest of which I will term “cognitive state”. MVPA analyses of cognitive states are those that investigate processes rather than specific representations (see section 1.7 for a discussion of this distinction). An example of this is evident in a study by Rissman et al. (2010) who demonstrated that it is possible to decode subjective mnemonic states (e.g. a feeling that a face is new or old) from whole-brain patterns of voxel activity. This MVPA analysis is not specific to any particular type of representation, but instead it speaks to the cognitive states/processes relating to recognition memory. It is worth noting that this type of information is not uniquely accessible to MVPA analyses, as the mass-univariate approach was originally developed in order to differentiate such neural processes, although MVPA may offer a more sensitive measure of information in some circumstances.

Stimulus categories, such as faces or places (or nouns and verbs) constitute another important type of information, and MVPA has been useful for elucidating the neural representation of such categories. Again, this level of information is available to mass-univariate analyses, as exemplified by

studies investigating the “fusiform face area” and “parahippocampal place area” (Kanwisher et al., 1997; Epstein and Kanwisher, 1998). However, as originally demonstrated by Haxby et al. (2001), there may be residual information about non-preferred categories within each of these regions when analysed with MVPA, which demonstrates that both approaches are necessary for a full understanding of the neural representation of categories.

The most subtle type of information is that pertaining to individual stimuli, or individual internal representations (e.g. a specific memory). This is the only type of information where mass-univariate analyses are not able to provide any useful information (although see earlier section on adaptation analysis), and where MVPA becomes essential. This is because it is not usually possible to find a regional difference in overall activation between e.g. two individual scenes or faces, using a mass-univariate analysis. A comparison of two MVPA studies of the MTL illustrates the different (but complementary) interpretations that can be made from investigating different levels of representation. Diana et al. (2008) used MVPA to investigate the representation of various stimulus categories (including scenes) in both the posterior parahippocampal gyrus and the hippocampus. They found evidence of category-level information within the former but not the latter. Bonnici et al. (2012) investigated the representation of individual scenes within the MTL, and found evidence for scene representations within both the parahippocampal cortex and the hippocampus. This suggests the hippocampus may contain more distinct representations of individual scenes which are not organised in a category-specific fashion, thereby allowing successful item decoding but not category

decoding. All four decoding experiments described in this thesis involve the investigation of item-level information, as I am particularly interested in the neural representation of individual episodic memories.

2.8 Segmentation of regions of interest

For all four decoding studies I was primarily interested in investigating episodic representations in the hippocampus and surrounding MTL structures. I therefore used a region-of-interest approach for each analysis. For the majority of studies the regions of interest were manually segmented using structural MR images, and the specific protocols used are included in the methods section of each experimental chapter. The only exception to this was Experiment 1, where I used a standard automated segmentation procedure called FreeSurfer (Fischl et al., 2002, 2004) in order to create the regions of interest. FreeSurfer uses two approaches to segmentation – one is designed for the cortex, and the other for the subcortical structures. The cortical "pipeline" involves flattening the cortex, and registering it to a spherical atlas, while the subcortical pipeline uses high-dimensional warping to normalise the structural image to a standard template. In each case, segmentation is then achieved through a probabilistic labelling system which assigns each voxel to one of the possible regions (Fischl et al., 2002, 2004). FreeSurfer segmentation performance has been shown to have a reasonable degree of accuracy compared to manual segmentation in some regions. However, the anatomical definitions of two key regions (the entorhinal cortex and posterior parahippocampal cortex) were, in my opinion, not optimal. I subsequently manually adjusted these two ROIs for

each subject following FreeSurfer segmentation, using the guidelines of Insausti et al. (1998). For increased accuracy, I therefore decided to manually segment all regions of interest for each subject in subsequent experiments.

2.9 Dynamic Causal Modelling

So far I have focussed on methods which aim to assess activation or information within specific anatomical regions. However, another important source of information may be present within the task-dependent connectivity between different regions. In other words, if two regions increase in connection strength during a particular task, this provides us with additional important information about the neural processing involved in that task. While this approach is not the focus of my thesis, the final experiment (Chapter 7) did make use of a particular connectivity method known as Dynamic Causal Modelling (DCM) in order to investigate the interaction between the hippocampus and cortical regions.

DCM is a Bayesian model comparison method which involves creating various plausible models of the task-dependent effective connectivity between pre-specified neural regions (Friston et al., 2003; Stephan et al., 2010). Once fitted, the evidence associated with each model can be compared in order to determine which is the most likely (or “winning”) model. More specifically, DCM models the neural dynamics of a specified system of interacting brain regions by representating the population activity at the neural level with a single state variable for each region (Friston et al.,

2003). The change in this state vector X in time is modelled as a bilinear differential equation, which for a single input U can be written as:

$$dx/dt = (a + Ub)X + Cu$$

In this equation, the A matrix represents the intrinsic connectivity between the given regions (which is defined as the average connectivity over the experiment), the B matrix represents the modulatory effect of experimental variables on these connections, and the C matrix provides the driving inputs to the system. For example, let's assume we have a task involving simple visual gratings presented either with explicit attention or without. The model involves two regions, the primary visual cortex (V1), and posterior parietal cortex (PPC). We assume that the two regions are connected bidirectionally (the A matrix), and we hypothesise that the connection from PPC to V1 will be modulated in the presence of attention (the B matrix). Finally, we assume that the driving input will be to the earliest visual processing region included in the system, which is clearly V1 in this case (the C matrix). DCM combines this model of neural dynamics with a forward haemodynamic model describing how the neural population activity induces changes in the BOLD signal (Stephan et al., 2007). The full model is then fitted to the data, producing a log model evidence for that model, which takes into account both the accuracy and the complexity of the model. The purpose of DCM analysis as a whole is to compare multiple plausible models of the neural dynamics in order to determine whether there is any significant evidence for one specific model (or one family of models - Stephan et al., 2010). DCM is implemented within SPM, and is now considered a standard method for assessing neural connectivity.

3 Chapter 3

Decoding individual episodic memory traces in the human hippocampus

Precis

As described in Chapter 1, the principal aim of this thesis is to find and deploy a new means of investigating information at the level of individual episodic memories. This will, I believe, facilitate our understanding of the neurobiological basis of individual episodic memory traces within the hippocampus, and provide novel empirical data to inform some of the major questions and debates in the episodic memory literature. In this chapter I describe Experiment 1, where I tested whether it is possible to decode individual rich episodic memories solely from patterns of BOLD activity across voxels in the human hippocampus, using high resolution fMRI and MVPA methods.

3.1 Introduction

The search for the elusive engram, or memory trace, in the brain has been an ongoing endeavour in neuroscience for nearly a century (Semon, 1923; Lashley, 1950; Dudai, 2004). Although the biological existence of such engrams coding for memories is widely accepted, the precise mechanisms, locations and even the nature of the engram itself, in light of processes such as reconsolidation (Nadel and Land, 2000; Dudai, 2004), is the subject of much debate. Clearly the components of a complex multi-modal memory, such as a rich episodic memory, are likely to be widely distributed throughout the cortex (Wheeler et al., 2000). These components on their own are not sufficient, however. Something must bind the disparate elements of a recent episodic memory together to allow the relevant neural

representations to co-activate thus facilitating recollection (Treves and Rolls, 1994; McClelland et al., 1995; Shimamura and Wickens, 2009; Rolls, 2010; O'Reilly et al., 2011). Marr (1971) proposed that the hippocampus provides this function by storing a memory 'index', a distilled representation containing the essence of the memory which is synaptically linked to the full representation stored in the neocortex. The hippocampus is ideally suited for multi-modal binding, given its purported location at the top of the sensory cortical hierarchy and widely acknowledged role in supporting episodic memory (Andersen et al., 2006).

Precisely how the hippocampus codes for episodic memories, however, is still unknown. This is because tracking an individual episodic memory in terms of the activity of the many thousands of hippocampal neurons that support it remains a substantial challenge (Gelbard-Sagiv et al., 2008; Hassabis et al., 2009), complicated further by the possibility that episodic memories might be uniquely human (Tulving, 2002; Suddendorf and Busby, 2003). MVPA techniques applied to human fMRI data (Haynes and Rees, 2006; Norman et al., 2006) may offer a means to bridge the gap between recordings from single neurons and examining episodic memory across large populations of neurons in humans. As described in the previous chapters, MVPA assesses local patterns of information across voxels, permitting the differentiation of distinct perceptual and mental states in a manner not possible using conventional univariate fMRI analyses (Haynes and Rees, 2006; Norman et al., 2006).

In a recent study, MVPA was used to decode spatial information and predict the location of participants in a virtual reality environment from the pattern of fMRI signals across voxels in the human hippocampus (Hassabis et al., 2009). Here, using high spatial resolution fMRI, I investigated whether it would be possible to predict which specific recent episodic memory a participant was recalling solely on the basis of the fMRI BOLD activity patterns across voxels in the hippocampus, thus potentially distinguishing specific memory traces.

Given that the entorhinal and posterior parahippocampal cortices are both major input pathways to the hippocampus (Amaral, 1999), I also investigated the episodic representations within each of these regions. This then allowed the consideration of decoding performance across the three MTL regions in order to infer the relative strength of episodic information in each region. If the hippocampus is the major locus of the engram, as hypothesised by many (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011), then we would expect decoding performance to be greater in the hippocampus than elsewhere in the MTL.

3.2 Methods

3.2.1 Participants

Ten healthy right-handed participants (six female) took part in the experiment (mean age 21.1 years, SD 1.8, range 18–24). All had normal or corrected-to-normal vision and gave informed written consent to participation in accordance with the local research ethics committee.

3.2.2 Pre-scan training

During a pre-scan training period, participants viewed three film clips of everyday events. Each clip was 7s long and featured a woman (a different woman in each clip) carrying out a short series of actions. The films were shot outdoors in three different urban settings. These stimuli ensured that memories would be episodic-like in nature, and that all participants recalled the same set of memories. One clip featured a woman taking a letter out of her handbag, posting it in a post box, and then walking off. Another clip featured a woman taking a drink from a disposable coffee cup, putting the cup in a rubbish bin, and then walking off (see Figure 12A). The final clip featured a woman picking up a bicycle that was leaning against some railings, adjusting her helmet and walking off with the bicycle. The participants saw each clip 15 times, and practised vividly recalling them. A further consideration was the length of time it took to recall the memory of a clip. As each memory would be recalled multiple times in the scanning session (on average 17 times), it was important that the temporal duration of the recall period was similar on each occasion. This temporal dimension was

therefore emphasised during training, and feedback was provided on the timing accuracy on each practice trial. This extensive training ensured that the duration of recall was consistent for each memory and across the three memories.

3.2.3 Task

There were two experimental conditions during scanning. The first involved a cued recall task where on each trial the participant was presented with a cue indicating which of the three film events they were required to recall (see Figure 12B). Following this, an instruction appeared on the screen indicating that the participant should close their eyes and vividly recall the cued memory. Participants were instructed not to begin the recall process until this instruction appeared, and were trained on this procedure in the pre-scan session. I also included a check that the participants were concentrating, and to make sure that the recall approximated the original 7s length of a clip. The participant had to press a button (using a scanner-compatible button-box) when they had finished recalling the clip. If the button was pushed too soon (<6s) or they failed to push it within 10s then the participant would hear a tone, and a message would appear for 1.5s indicating that their recall had been too fast or too slow. Any such trials were excluded from the subsequent analysis. If the participant pressed the button between 6-10s, a fixation cross appeared onscreen for 1.5s. Participants were trained to open their eyes as soon as they had pressed the button or if they heard a tone. Following this, the participant was required to provide ratings about the preceding recall trial using the five-key button-box. Firstly, they rated how vivid the preceding recall trial was (scale: 1 – 5, where 1 was not vivid at all,

and 5 was extremely vivid). Secondly, they rated how accurately the recalled memory reflected the actual film clip (scale: 1 – 5, where 1 was not accurate at all, and 5 was extremely accurate). Any trials where a participant recorded a rating of less than 3 were excluded from the subsequent analysis. Following the ratings, participants rested for 4s before starting the next trial. The cued recall condition contained a total of 21 trials, with seven trials of each memory, presented in a pseudo-random order, whilst ensuring that the same memory was not repeated twice or more in a row.

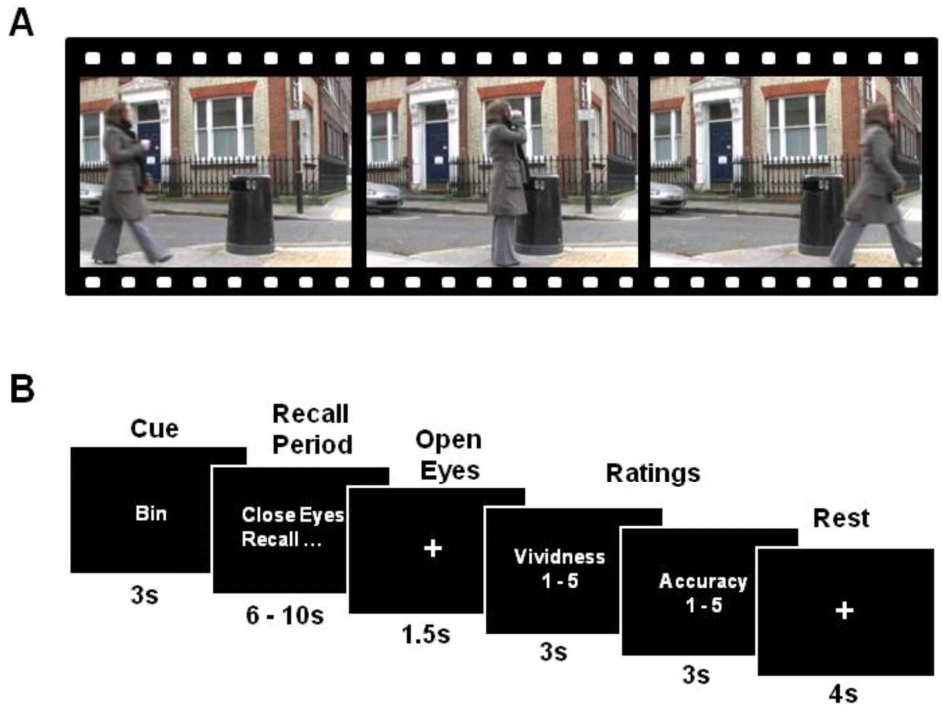

Figure 12. Experimental design. *(A) Still photographs taken from one of the film clips viewed during pre-scan training. The clip depicted a woman taking a drink from a disposable coffee cup and then putting it in a rubbish bin. (B) Timeline of an example cued recall trial during fMRI scanning.*

The second condition was a free recall task, where the participant was allowed to decide which of the three episodes they would recall on each trial. Here, the cue period was replaced with a decision period, during which the participant decided which of the three memories they would subsequently recall. The same procedure as cued recalled was then followed, with the addition that after the recall period, participants were required to indicate via the button-box which of the three memories they had just recalled. Ratings of vividness and accuracy were again taken for each trial. This condition included a total of thirty trials, and participants were instructed to sample from the three memories, while avoiding the recall of the same memory twice in a row. In order to ensure that participants did not sample the memories in a predictable order, I calculated the probability that each memory was followed by each other memory, created a set of pair-wise statistical dependencies for each participant. These are displayed in Table 1. Both experimental conditions were scanned in a single functional run, starting with the cued recall condition, with a thirty second rest period before the free recall condition.

After the scanning session, participants answered a debriefing questionnaire, which was designed to assess aspects of their memory recall. They were asked to provide ratings (on a scale of 1 – 5, low - high) for each of the three memories based on the average response across all trials during scanning for the following:

How hard did you find it to vividly recall this event?

How emotional did this event make you feel?

How much did this event make you think about a real memory from your own life?

How much did this event make you think about yourself?

How much did you find yourself thinking about some sort of background story behind the event?

How much did you find yourself trying to take the perspective of the person in these events?

3.2.4 Image acquisition

All functional images were acquired using the high-resolution fMRI sequence that I described in Chapter 2. Field maps were acquired for distortion correction. T1-weighted MDEFT whole-brain structural scans were acquired for each participant after the main scanning session.

	AB	AC	BA	BC	CA	CB
P1	0.45	0.55	0.56	0.44	0.67	0.33
P2	0.29	0.71	0.44	0.56	0.3	0.7
P3	0.63	0.38	0.55	0.45	0.25	0.75
P4	0.88	0.13	0.22	0.78	0.75	0.25
P5	0.44	0.56	0.5	0.5	0.56	0.44
P6	0.44	0.56	0.6	0.4	0.33	0.67
P7	0.44	0.56	0.5	0.5	0.4	0.6
P8	1	0	0.1	0.9	0.88	0.13
P9	0.73	0.27	0.64	0.36	0.43	0.57
P10	0.5	0.5	0.67	0.33	0.86	0.14
Mean	**0.58**	**0.42**	**0.48**	**0.52**	**0.54**	**0.46**
SD	**0.22**	**0.22**	**0.18**	**0.18**	**0.23**	**0.23**

Table 1. Free Recall statistical dependencies. *Pairwise statistical dependencies displayed by each participant (P1 – P10) during the Free Recall condition, along with group mean and standard deviation. Column 2 displays the probability that the recall of memory A was followed by the recall of memory B, column 3 displays the probability of memory C following memory A, and so on. Note that participants were explicitly instructed not to recall the same memory twice in a row; therefore the probably of each memory being followed by itself is zero. Some participants display strong dependencies, but the group as a whole was well balanced, and none of the group dependencies was significantly different from chance (50%).*

3.2.5 Univariate analysis

A standard mass univariate statistical analysis was performed using SPM8. The first six EPI volumes were discarded to allow for T1 equilibration effects (Frackowiak et al., 2004). Spatial pre-processing comprised realignment and normalisation to a standard EPI template in Montreal Neurological Institute (MNI) space, and smoothing using a Gaussian kernel with FWHM of 8mm. After pre-processing, statistical analysis was

performed using the general linear model. Each of the three memories was modelled as a separate regressor, where the recall period of each trial was modelled as a boxcar function and convolved with the canonical hemodynamic response function. Participant-specific movement parameters were included as regressors of no interest. Participant-specific parameter estimates pertaining to each regressor (betas) were calculated for each voxel. These parameter estimates were entered into a second level random-effects analysis using a one-way ANOVA, with the three memory regressors as the three factors in the ANOVA. Given my *a priori* interest in the medial temporal lobes, a significance threshold of $p<0.001$, uncorrected for multiple comparisons, was employed for voxels within this region. A significance threshold of $p<0.05$ corrected for family-wise errors was employed for voxels elsewhere in the partial volume.

3.2.6 Image pre-processing for multivariate analysis

T1-weighted structural images were put through the FreeSurfer (Fischl et al., 2002, 2004) processing pipeline in order to generate a set of anatomical regions of interest (ROIs). FreeSurfer automatically assigns an anatomical label to each voxel based on a probabilistic atlas, and the technique has been shown to be comparable in accuracy to manual labelling (Fischl et al., 2002, 2004). This generated a set of hippocampus (HC), entorhinal cortex (EC), and posterior parahippocampal cortex (PHC) masks for each participant. The anterior and posterior boundaries of the entorhinal and parahippocampal masks were altered manually where necessary to ensure that they were in line with the anatomical guidelines set out by Insausti et al. (1998).

The first six EPI volumes were discarded to allow for T1 equilibration effects (Frackowiak et al., 2004). The remaining EPI images were then realigned to correct for motion effects, and minimally smoothed with a 3mm FWHM Gaussian kernel. This minimal smoothing was included in order to reduce noise from potential residual misalignments between scans, while still ensuring that information was present at a fine-grained spatial resolution. A linear detrend was run on the images to remove any noise due to scanner drift (LaConte et al., 2005). Next the data were convolved with the canonical haemodynamic response function (HRF) to increase the signal-to-noise ratio (Frackowiak et al., 2004). This HRF convolution effectively doubled the natural BOLD signal delay, giving a total delay of approximately 12s. To compensate for this delay, all onset times were shifted forward in time by three volumes, yielding the best approximation to the 12s delay given a TR of 3.5s and rounding to the nearest volume (Haynes and Rees, 2006). Functional volumes were extracted from the vivid recall period of each trial, leading to a total of between two and four functional volumes per trial, depending on the precise start-time and length of the recall period in each case.

3.2.7 MVPA classification

In order to assess the degree of episodic information contained within MTL structures, I used a two-step procedure incorporating first feature selection and then final multivariate classification (Guyon and Elisseeff, 2003). As explained in Chapter 2, the purpose of feature selection is to reduce the set of features (in this case, voxels) in a dataset to those most likely to carry relevant information. The particular feature selection strategy employed was

a multivariate searchlight strategy (Kriegeskorte et al., 2006), which assesses the local pattern of information surrounding each voxel in turn (see feature selection section below for more details). See Figure 13 for an overview of the entire classification procedure. This figure includes elements already provided in Figure 11 (Chapter 2), but I feel it is worth reiterating them again here in the context of their specific experimental deployment.

The overall classification procedure involved splitting the imaging data into two segments: a "training" set used to train a linear support vector machine (SVM) with fixed regularization hyperparameter C = 1, in order to identify response patterns related to the memories being discriminated, and a "test" set used to independently test the classification performance (Duda et al., 2001). Prior to multivariate classification, feature selection was performed on the data from the training set (thereby ensuring that this step was fully independent from final classification, which is critical for avoiding "double-dipping" - Kriegeskorte et al. 2009; see also Chapter 2). This step produced a subset of voxels within the ROI that contained the greatest degree of episodic decoding information within the training dataset. Using this voxel subset, the SVM classifier was trained to discriminate between the three memories using the "training" image dataset, and tested on the independent "test" dataset **(**see Figure 13). The classification was performed with a SVM by using the LIBSVM implementation (Chang and Lin, 2011). I used a standard k-fold cross-validation testing regime (Duda et al., 2001) wherein k equalled the number of experimental trials, with the data from each trial set aside in turn as the test data, and the remaining data used as the training set.

This therefore generated k sets of SVM training and test sets which produced an overall classification accuracy from the proportion of correct classification “guesses” across all k folds of the cross-validation.

Note that standard SVMs are binary classifiers that operate on two-class discrimination problems, whereas my dataset involved a three-class problem. The SVM can, however, be arbitrarily extended to work in cases where there are more than two classes. Typically this is done by reducing the single multiclass problem into multiple binary classification problems that can be solved separately and then recombined to provide the final class prediction (Allwein et al., 2000). I used the well-established Error Correcting Output Codes approach (Dietterich and Bakiri, 1994) to assign a unique binary string to each of the three classes. The length of the binary string corresponds to the number of binary classifiers performed. As there are 3 possible pair-wise comparisons that can be made between the three memories, the unique binary string “codewords” were 3 bits in length. The 3 possible binary classifications were performed in each case, and their outputs combined into a 3-bit output code, with each bit representing the output from a single binary classifier. These output codes were then compared against all 3 of the pre-assigned class codewords to determine the final predicted class. This was achieved by computing the Hamming distance (Hamming, 1950) (i.e. the number of bits which differ between two binary strings) between the output code and the class codewords to find the closest fit. The memory represented by this codeword was then chosen as the output of the classification.

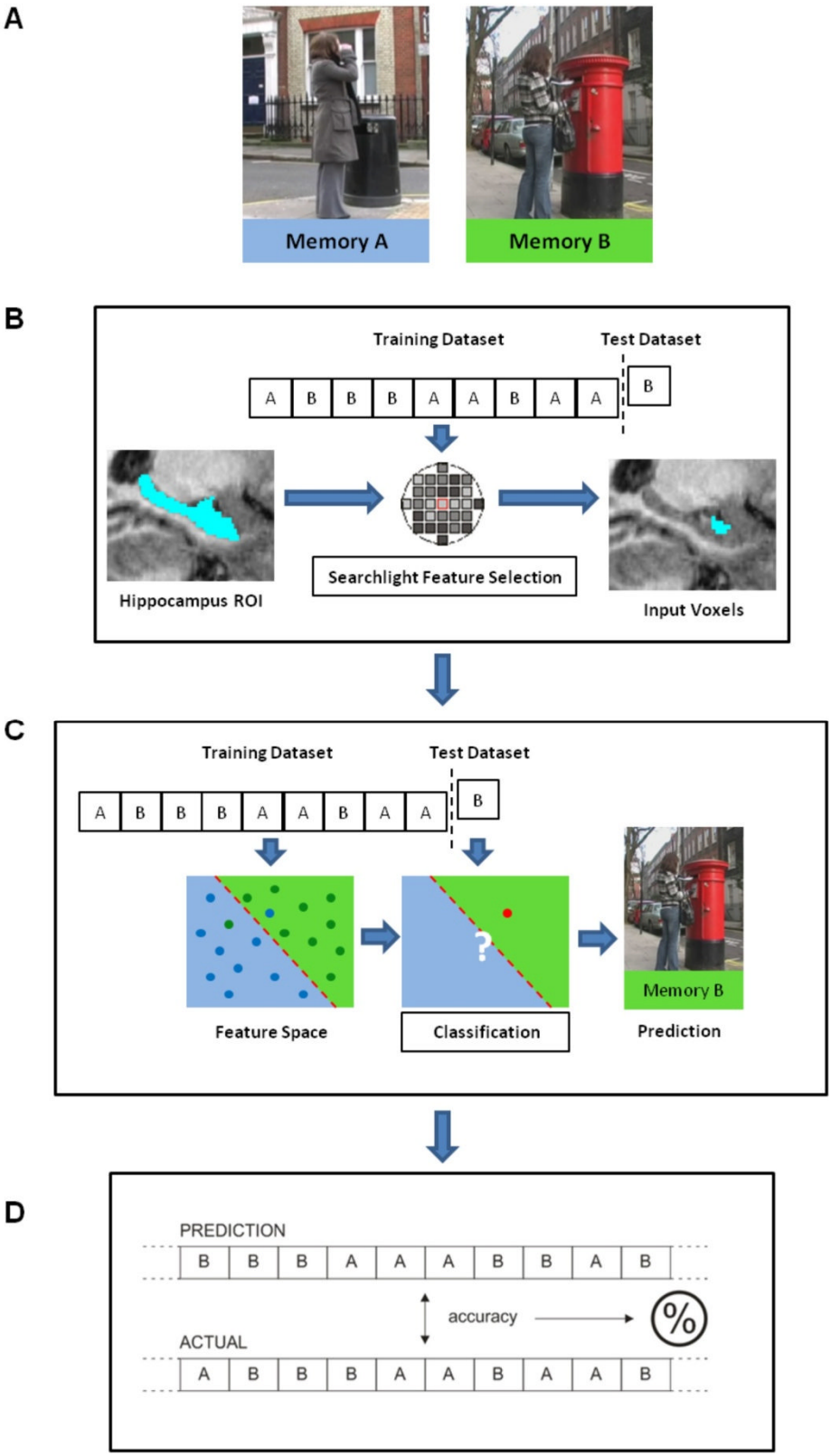

Figure 13. The overall MVPA classification procedure. *(A) For simplicity I demonstrate the procedure when classifying two distinct episodic memories, while in reality I classified three memories. In this case Memory A involved a woman sipping from a disposable coffee cup and putting into a rubbish bin and Memory B involved a woman posting a letter into a postbox.*

(B) Volumes acquired during the recall period of each trial were extracted, and labelled as memories A or B. The full dataset was split into a "training" set and a "test" set, where the test set was the data from a single experimental trial. Using the training set, searchlight feature selection was applied to the voxels within the region of interest (ROI), in this example the hippocampus. This resulted in a reduced set of voxels which carried the most information. (C) Using the reduced voxel set, a classifier was trained to differentiate memories A and B using the training dataset, and then tested using the fully independent test set. In this case the test trial was classified as Memory B, which was a correct prediction. (D) A standard k-fold cross-validation testing regime was implemented, ensuring that all trials were used once as the test data set. This cross-validation therefore yielded a predicted label for every data trial in the analysis, which was then compared to the real labels to produce and overall prediction accuracy value.

3.2.8 Feature selection

Feature selection was implemented using a multivariate searchlight strategy (Kriegeskorte et al., 2006), which examines the information in the local spatial patterns surrounding each voxel within the search space. Thus, for each voxel within the chosen anatomical ROI, I investigated whether its local environment contained information that would allow accurate decoding of the three memories. For a given voxel, I first defined a small sphere with a radius of three voxels centred on the given voxel. This radius was chosen because a previous demonstration of hippocampal decoding using the searchlight method used radius three (Hassabis et al., 2009). Note that the "spheres" were restricted so that only voxels falling within the given region of interest were included. Therefore the shape of the "sphere", and the number of voxels within it varied depending on the proximity to the ROI's borders.

A linear SVM was then used in order to assess how much episodic information was encoded in these local pattern vectors. This was achieved

by splitting the feature selection data-set into a training set and a test set (again it is important to note that all of the data used in this feature selection step is derived from the *training* set of the overall classification procedure, and therefore is fully independent of the final classification). The training set was then used to train a SVM classifier using the LIBSVM implementation and a fixed regularization hyperparameter of C = 1. I used a standard k-fold cross-validation testing regime (Duda et al., 2001) wherein k equalled the number of experimental trials minus one (as one trial is already removed for use as the overall testing set – see above), with the data from each trial set aside in turn as the test data, and the remaining data used as the training set. This therefore generated k sets of SVM training and test sets which produced an overall classification accuracy from the proportion of correct classification "guesses" across all k folds of the cross-validation. This procedure was repeated for each searchlight sphere, thus generating a percentage accuracy value for every single voxel within the search space.

The searchlight analysis described above therefore produces an "accuracy map" of the given ROI, with an accuracy value at each voxel representing the amount of decoding information contained within the searchlight sphere surrounding that voxel. This allows us to perform feature selection by selecting searchlight spheres with high accuracy values. In this case, the searchlight with the maximal accuracy value was chosen as the output of feature selection. In cases where more than one searchlight carried the maximal accuracy value, all voxels from all the maximal searchlight spheres were included as the feature selection output. See Figure 14 for an illustration of the feature selection process.

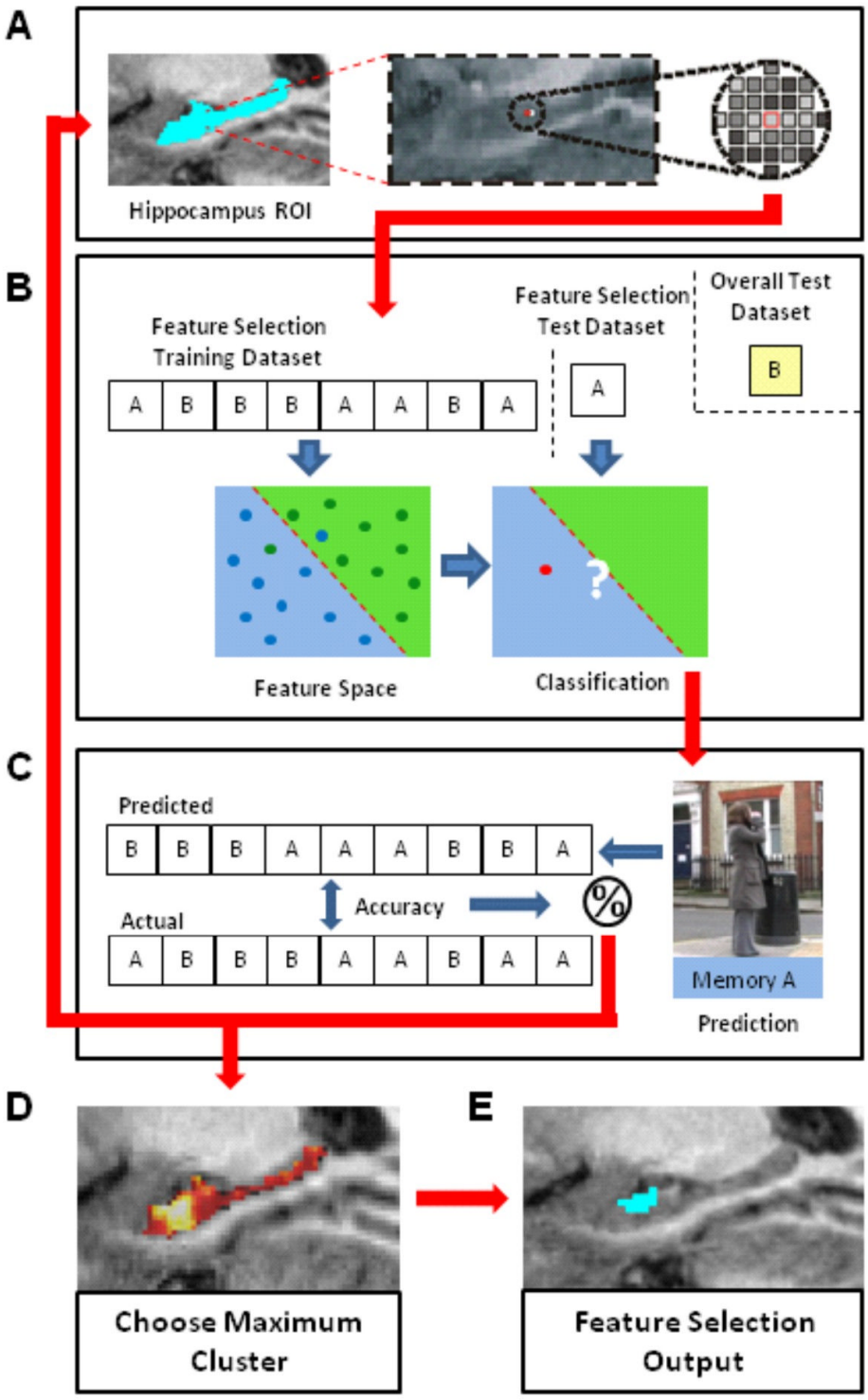

Figure 14. The searchlight feature selection procedure. *(A) The searchlight analysis stepped through every single voxel in the search space, which was defined by an anatomical ROI, in this example the hippocampus. For each voxel (example shown in red), a spherical cluster (radius 3 voxels) of 99 voxels was extracted from around this central voxel. (B) Once the overall test dataset had been removed, the remaining feature selection data was separated into a training set and a test set (which was the data from a single experimental trial). Using the voxel cluster from the searchlight, a*

classifier was trained to differentiate memories A and B using the training data, and then tested using the independent feature selection test data. (C) In this case the test trial was classified as Memory A, which was a correct prediction. A standard k-fold cross-validation testing regime was implemented, ensuring that all data trials were used once as the test data set. This cross-validation therefore yielded a predicted label for every trial in the analysis, which was then compared to the real labels to produce an overall prediction accuracy value. The whole procedure was then repeated for every single voxel within the search space. (D) This created an "accuracy map" of the whole ROI, with an accuracy value at each voxel representing the amount of information contained within the searchlight sphere surrounding that voxel. Here the accuracy values for each voxel are displayed in a heatmap. (E) For the feature selection output, the searchlight cluster with the highest accuracy value, and therefore greatest amount of information, was chosen and it is this voxel set that was fed into the overall classification analysis.

3.2.9 Information maps

The multivariate pattern analysis technique uses a feature selection procedure in order to select subsets of voxels more likely to carry information. This means that for each fold of the k-fold cross-validation, a different subset of voxels is selected. In order to visualise the voxels selected during feature selection, an "information map" was created by simply finding all voxel sets which produced above-chance accuracy on that particular cross-validation fold. These voxel sets were added together to form a single binary mask.

3.2.10 Overlap analysis

To investigate the consistency of location of decoding across participants, the individual hippocampal information maps were normalized using the FreeSurfer high-dimensional warps previously generated during creation of the anatomical ROIs. The ten information maps could then simply be added together to form a frequency heatmap. Assuming that the voxel location of

individual information maps follows a binomial distribution, the likelihood of finding the same voxel by chance N times out of 10 was assessed for each voxel frequency value N.

3.2.11 Temporal dependencies: control analysis

For any classification study, it is important to ensure that the data used for testing is independent of that used for training. In the current study the temporal gap between each recall period was at least 10s, which should ensure that the testing and training data are relatively independent. However, to test this assumption, I conducted a control analysis where I increased the temporal gap between the testing and training data. If residual temporal dependencies were affecting the results, then this increased temporal gap should significantly impair classification performance. This analysis was identical to the main analysis, but on each fold of the k-fold cross-validation, the trials that were temporally adjacent to the testing trial (trials k-1 and k+1) were excluded from both the feature selection data and the training data. This effectively increased the temporal gap between training and testing data to at least 26s.

3.3 Results

3.3.1 Behavioural Results

Table 2 displays a summary of participants' behavioural performance during scanning. Note that these summary statistics were derived after exclusion of low rating trials and trials that were too long or short (see Methods section). For each variable a repeated measures ANOVA was applied to determine if there were consistent differences between the three memories, and none of these analyses found any significant results (number of trials: $F = 0.74$, $p = 0.49$; recall length: $F = 1.51$, $p = 0.25$; vividness: $F = 0.003$, $p = 1$; accuracy: $F = 0.04$, $p = 0.96$).

	Postbox	Bicycle	Bin
No. of trials	14.5 (1.27)	14.6 (2.88)	13.7 (2.63)
Recall Length (s)	7.74 (0.2)	7.98 (0.51)	7.94 (0.43)
Vividness (1-5)	4.01 (0.56)	4.01 (0.55)	4.02 (0.61)
Accuracy (1-5)	4.01 (0.53)	4.01 (0.53)	3.99 (0.63)

Table 2. Behavioural results. *Means for number of trials, length of recall period, vividness, and accuracy ratings are displayed for each of the three memories collapsed across both the cued and free recall tasks. Standard deviations are displayed in parentheses. These summary statistics were derived after exclusion of low rating trials and trials that were too long or too short.*

The mean debrief ratings for each episode are displayed in Table 3. For each variable, the participants provided a rating on a scale of 1 – 5 (low – high). A repeated-measures ANOVA was used to test for significant differences

between the three episodes for each variable, and none of these analyses found any significant effects using a Bonferroni-corrected threshold of $p < 0.0083$ for multiple comparisons (Difficulty: $F = 0.7$, $p = 0.51$; Emotionality: $F = 0$, $p = 1$; Similarity to real memory: $F = 0.057$, $p = 0.95$; Thinking about oneself: $F = 0.41$, $p = 0.67$; Perspective-taking: $F = 3.62$, $p = 0.048$; Background story: $F = 1.22$, $p = 0.32$). Additionally, participants were asked whether they recognised the person or location featured in each event, and to give a rating (1-5, low-high) of their general attention during scanning. No participants recognised the people or places. The mean rating of attention was 4.1 (SD 0.57). Thus, the behavioural results demonstrate that the three episodes were matched across a variety of different measures, rendering it unlikely that any such extraneous factors could affect the MVPA analyses.

	Postbox	Bicycle	Bin
Difficulty	2.2 (0.79)	2.3 (0.95)	1.9 (1.1)
Emotionality	1 (0)	1 (0)	1.2 (0.42)
Similarity to real memory	1.7 (0.82)	1.7 (0.82)	1.6 (0.84)
Thinking about self	1.5 (0.71)	1.4 (0.7)	1.7 (0.82)
Perspective-taking	1.9 (0.88)	1.2 (0.42)	1.8 (1.03)
Background story	2 (1.49)	1.3 (0.48)	1.6 (0.97)

***Table 3. Debriefing questionnaire results**. Mean ratings are provided for each of the three memories for each questionnaire item, with standard deviations in parentheses. Participants were asked to provide ratings on a scale of 1 – 5 (low – high).*

3.3.2 Univariate Results

Using the mass-univariate approach, no significant differences in activity were apparent anywhere in the brain. These null univariate results were expected because the conventional univariate approach works by measuring gross voxel activity differences between conditions. With all conditions involving identical processes (episodic retrieval), it is not surprising that this method did not reveal any significant differences, hence the advantage of using a multivariate approach.

3.3.3 MVPA Results

No differences in decoding accuracy were found between the hemispheres for each MTL region, and this was the case for cued and free recall (all p values > 0.1). Decoding accuracies were therefore pooled across hemisphere for all subsequent analyses. I next tested for differences in decoding accuracy between the two retrieval conditions (cued and free recall), and found no significant differences in any of the three MTL regions (HC: $t = 1.42$, $p = 0.19$; EC: $t = 0.63$, $p = 0.54$; PHC: $t = 0.57$, $p = 0.58$). This demonstrates that the decoding accuracy does not depend on the specific retrieval mode. Therefore, for all subsequent analyses, the data were collapsed across both conditions in order to investigate patterns of information that hold across different retrieval modes.

Overall, the MVPA decoding analysis produced significant results across all three MTL regions (see Figure 15). This demonstrates that it is possible to predict which specific episodic memory was being recalled solely from the

pattern of fMRI BOLD signals across voxels in the hippocampus or neighbouring MTL cortex. However, a one-way repeated measures ANOVA revealed that the decoding accuracies were not equal across all three MTL regions ($F = 4.42$, $p = 0.027$). Post-hoc paired t-tests clearly demonstrated that this effect was driven by hippocampal accuracy being significant greater than either the entorhinal ($t = 2.48$, $p = 0.035$), or parahippocampal cortex ($t = 2.29$, $p = 0.048$). This result suggests that there is a significantly greater level of episodic information within the hippocampus than in surrounding MTL cortex, which is what we would expect given the functional and anatomical hierarchy of the MTL (Treves and Rolls, 1994; McClelland et al., 1995; Amaral, 1999; Rolls, 2010; O'Reilly et al., 2011).

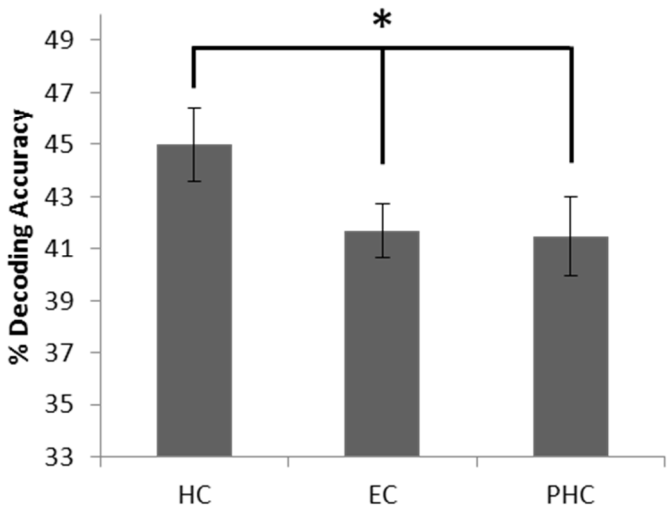

Figure 15. MVPA decoding results. *Mean decoding accuracy results with standard errors for the hippocampus (HC), entorhinal cortex (EC), and posterior hippocampal cortex (PHC). Percentage accuracy values are on the vertical axis; 0.33 represents chance level performance. All three regions were significantly above chance level performance, with HC accuracy significantly greater than both EC and PHC.*

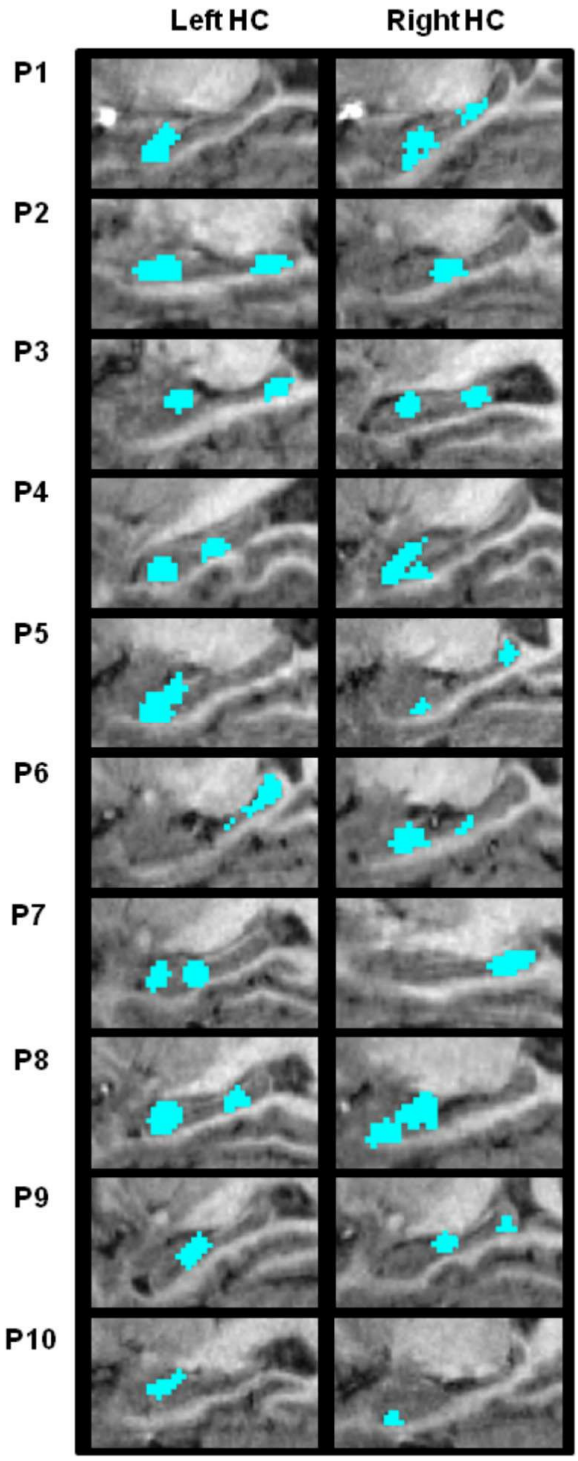

Figure 16. Hippocampal information maps. *The information maps in the left and right hippocampi are shown for the ten participants (P1-10) on zoomed-in sagittal sections of the medial temporal lobes taken from each participant's structural MRI scan. Each map represents the set of voxels carrying the most episodic information within the hippocampus. Note the consistencies across subjects, particularly in the anterior hippocampus.*

3.3.4 Information maps

A priori it is not clear whether particular regions within the hippocampus should show a preference for coding individual episodic memories. A useful property of the feature selection method used in this analysis is that it produced a sub-set of voxels within a region of interest that carried the most episodic information (see Methods). I refer to this as the "information map" for that region, and the hippocampal information maps for all ten participants are displayed in Figure 16. An inspection of these maps suggests there may be consistencies across participants in the location of episodic information.

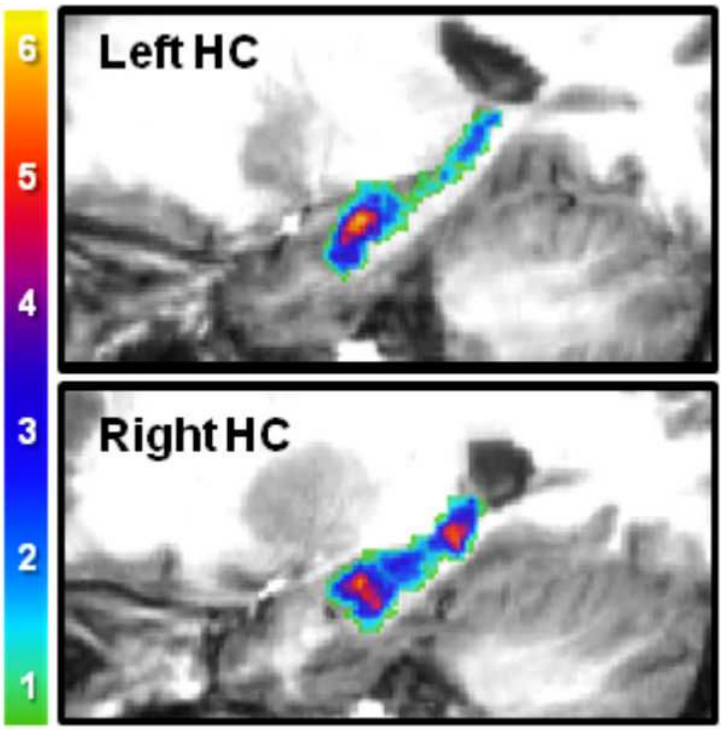

Figure 17. Information heatmaps*. Frequency heat-maps for the left and right hippocampi shown on zoomed-in sagittal sections from one of the participant's structural MRI scans chosen at random. Frequency scale is shown on the left. To determine statistical significance, the frequency value at each voxel was compared against the binomial distribution, and the peak regions in yellow and red all survive an uncorrected p<0.001 level of significance.*

To examine this further, the hippocampal information maps for all ten participants were transformed into standard stereotactic space, and added together to form a frequency heatmap (Figure 17). This heatmap clearly shows three peak regions of overlap, in bilateral anterior and right posterior hippocampus. All three peak regions in red and yellow are significant at a threshold of $p < 0.001$. This result demonstrates that episodic information is not randomly distributed across the hippocampus, but is instead concentrated within specific regions.

3.3.5 Temporal dependencies: control analysis

This control analysis was included in order to rule out any possible temporal "carry-over" of information between temporally adjacent trials (see Methods). This analysis produced significant decoding accuracies in all three MTL regions, with mean hippocampus accuracy of 44% ($p = 0.000001$; chance level = 33%), mean EC accuracy of 38.5% ($p = 0.009$), and mean PHC accuracy of 41% ($p = 0.0004$). A direct comparison of these new results with the original results did not find significant differences for any of the three MTL regions. These results demonstrate that the addition of a substantial temporal gap between testing and training data does not make any significant difference to the decoding performance, and I can therefore be confident that my training and testing data were independent.

3.3.6 Comparison of cued and free recall conditions

To ensure that the decoding results were based on information that was consistent across the two modes of retrieval, I performed a further control analysis. A searchlight classifier was applied to the hippocampi using only the free recall data. The maximal searchlight was found, and this set of voxels was then used to train on the free recall data and test on the cued recall data. Given the large reduction in training data that results from this procedure, one would expect a considerable loss in classifier sensitivity using this approach. Nevertheless, collapsing across both hippocampi there was a trend towards significant decoding (mean accuracy 35%, $p = 0.12$; chance = 33%) with a significant result in the right hippocampus (mean accuracy 38%, $p = 0.028$). These results demonstrate that the classifier is making use of common information across the different conditions, and does not rely on information that is specific to the mode of retrieval.

3.4 Discussion

In this experiment I have provided the first evidence that individual episodic memories can be differentiated based solely on patterns of BOLD activation across the human hippocampus and surrounding MTL cortex during episodic retrieval. This demonstrates that MVPA decoding, in combination with high-resolution fMRI, presents a viable new method for investigating episodic information at the level of individual memories in vivo and non-invasively. This is an important advance, as we currently have very little concrete evidence regarding the neural representation of specific episodic memories, despite the existence of detailed theoretical models (Marr, 1971;

Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Further use of this method may allow us to start mapping out the representational properties of episodic memories, and to provide empirical tests of existing theories.

Beyond the important future applications of this method, the specific results of this study provide some initial insights into the nature of episodic representations. First, the results imply that the neuronal traces of the memories were stable even over many re-activations. This is due to the fact that the MVPA decoder could only perform successfully if the patterns of activation for each memory were stable and consistent across the many retrieval trials of the experiment. While this is perhaps not a surprising result, it is nevertheless the first time that this property has been empirically demonstrated.

Second, the decoding accuracy was significantly greater in the hippocampus than either of the two MTL cortical regions. This suggests that the episodic representations were in some way more distinct or "stronger" within the hippocampus than in either the entorhinal or parahippocampal cortices. This result is entirely consistent with existing theoretical models of episodic memory which propose that information about individual stimuli, such as objects, people, and spatial contexts, are processed within the MTL cortex. These episodic "elements" are then passed into the hippocampus where they are bound into a single coherent episodic representation. Notably, the episodic representations within the hippocampus are proposed to be orthogonalized into distinct neural representations through the process of

"pattern separation" (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). These models would therefore explicitly predict that episodic representations ought to be more distinct and separable within the hippocampus than in the earlier, input regions in MTL cortex. As such, these results fit in well with the extent theoretical framework for episodic representation.

It is worth discussing one important caveat at this point – direct comparisons of the MVPA classification accuracy of different regions should be interpreted with caution. The reason for this is that there may be other variables besides the underlying neuronal information that could lead to differences in accuracy value (Diedrichsen et al., 2011). For instance, if the BOLD signal is noisier in one region than another, this is likely to lead to lower accuracy values in the noisier region, even if the underlying neuronal information is actually the same. Similarly, if two regions contain different numbers of voxels, then this could lead to differences in decoding accuracy. This can influence the results in either way, as larger regions could lead to increased accuracy (due to increased numbers of informative voxels in that region) or decreased accuracy (due to increased numbers of noisy voxels in that region). Thus, while the comparison of hippocampal versus MTL cortical decoding accuracies are suggestive, particularly given the fact that they are consistent with theoretical views of episodic representation, we cannot draw strong conclusions based on this method alone.

The third major result of this study was the demonstration that the episodic information was not randomly distributed across the hippocampus, but

instead was concentrated within bilateral anterior and right posterior regions. This suggests that there may be some functional topography within the hippocampus, with preferential episodic processing within these specific regions. Previous studies have noted potential functional and anatomical dissociations between the anterior and posterior hippocampus. For example, Kahn et al. (2008) demonstrated that the patterns of connectivity between posterior and anterior hippocampus and the rest of the brain were distinct in human subjects at rest, and finding which was replicated in a recent study by Poppenk and Moscovitch (2011). Similarly, evidence from behavioural, anatomical, and gene expression studies all support the idea that anterior and posterior hippocampi are at least partly dissociable in rodents as well (Moser and Moser, 1998; Fanselow and Dong, 2010). The right posterior hippocampus in particular is frequently associated with spatial processing (Maguire et al., 2000; Hassabis et al., 2009; Woollett and Maguire, 2011), and I speculate that this region may be particularly involved in the representation of spatial elements of the episodic memories in the current study. The robust loci in bilateral anterior hippocampal regions, on the other hand, are consistent with previous studies of autobiographical memory (Svoboda et al., 2006), and represent a clear target for future investigations.

In summary, I documented that traces of individual rich episodic memories are detectable and distinguishable in the human hippocampus. This demonstrates the viability of applying MVPA decoding techniques to the study of episodic representations, allowing us to directly access information about individual episodic memories in the human hippocampus in vivo and non-invasively. This offers exciting new opportunities to examine important

properties of episodic memory, and to start exploring the neurobiological basis of the “engram”, that I will pursue in the subsequent experimental chapters.

3.5 Clinical applications

Besides the important theoretical questions that this approach allows us to address, MVPA has potential clinical applications as well. As part of my PhD I collaborated on a feasibility study with my colleague Heidi Bonnici, and clinicians Meneka Sidhu and John Duncan at the National Hospital for Neurology and Neurosurgery. While this study does not form a core part of my thesis, I briefly mention it here to illustrate the potential applications of MVPA to clinical populations. The purpose of the study was to use the same design and analysis as described in Experiment 1 above in order to investigate the episodic information present within the hippocampi of ten patients with temporal lobe epilepsy and unilateral hippocampal sclerosis (9 with hippocampal sclerosis on the left and 1 on the right). The aim of the study was to ascertain whether it would be possible to decode memories from the pathological hippocampus (we hypothesised not) and, more importantly, whether the ‘intact’ hippocampus was able to support viable memory representations. As the patients were being considered for unilateral temporal lobectomy for the relief of their seizures, knowing whether the remaining hippocampus was functional or not could aid in the decision to operate.

We found that the decoding results in the non-sclerotic MTL were very similar to those described above, with significant decoding accuracy within the hippocampus and surrounding MTL cortex, and with hippocampus performing significantly better than the cortical regions (see Figure 18). The sclerotic hippocampi, on the other hand, showed a marked impairment in decoding accuracy, and did not achieve above-chance performance at the group level, suggesting that the residual tissue here was not contributing to the representation of episodic memories. Interestingly, the MTL cortical regions in the sclerotic hemisphere both produced significant decoding accuracies, and indeed the accuracies were no worse than the intact MTL. This suggests that the entorhinal and parahippocampal cortices may retain functionality despite the neighbouring sclerotic hippocampus.

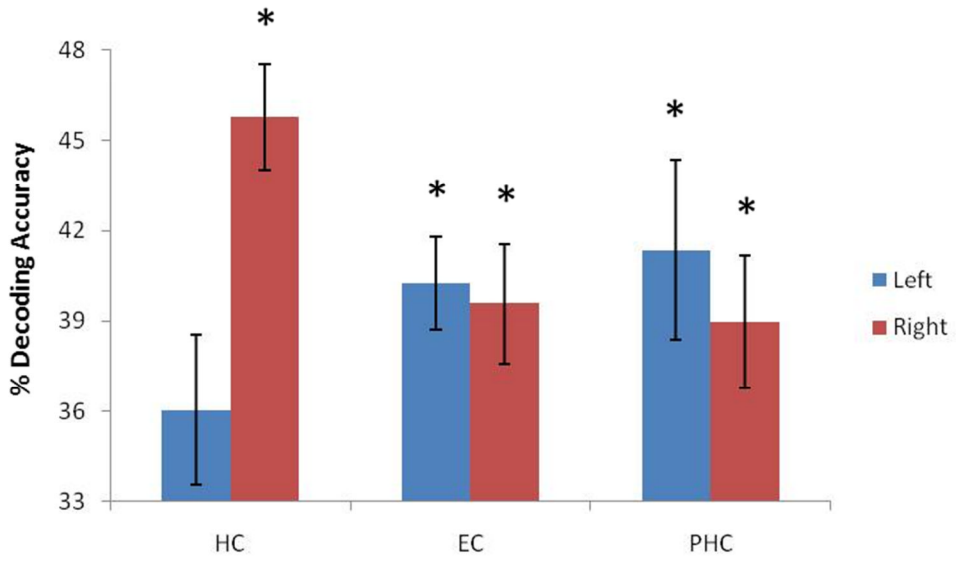

Figure 18. Episodic memory decoding in patients with unilateral hippocampal sclerosis and intractable epilepsy. *Data from the nine patients with left hippocampal sclerosis are shown here, with percentage accuracy on the y axis. The opposite pattern (i.e. left much better than right hippocampal decoding) was apparent in the patient with right hippocampal sclerosis. Asterisks indicate above-chance decoding accuracy. It is clear that the sclerotic left hippocampi did not support viable memory decoding, while the 'intact' right hippocampi had decoding accuracies in line with those of the healthy participants described in Experiment 1 above.*

These decoding results show that when the underlining neurons are dysfunctional, then the classifier cannot perform, confirming that MVPA is indexing neuronal activity. While preliminary and involving just 10 patients, the decoding results in patients with epilepsy mirrored their functional impairments. The ultimate goal of this line of research is to find a reliable tool that can predict neuropsychological outcome following surgery. If MVPA accuracy can predict whether there is sufficient healthy tissue to "take up the slack" following resection of the sclerotic tissue, this would prove to be an invaluable tool for clinical decision-making.

4 Chapter 4

Decoding recent and remote autobiographical memories

Precis

In the previous study I demonstrated that it is possible to decode individual episodic memories from the pattern of activity within the human hippocampus. In Experiment 2, I wanted to use this novel approach to shed new light on one of the major debates in the field of episodic memory – consolidation. As described in Chapter 1 and below, it is currently not clear whether the hippocampus is involved in the representation of remote episodic memories. Neither neuropsychological studies nor classical univariate fMRI analyses have fully resolved this debate. By using MVPA decoding to assess episodic memories at the level of individual memory traces, I hoped to shed some new light on the representation of remote episodes in the human hippocampus. In order to study remote memories, it was not possible to use the movie stimuli that I developed for the first experiment. Instead, in Experiment 2 the focus was on autobiographical memories (personally meaningful episodic memories) that were two weeks old, and autobiographical memories that were more than a decade old, to compare the representation of both types of memory in the hippocampus, and elsewhere.

This study was carried out in collaboration with my colleague Heidi Bonnici. We both contributed to the study design, Heidi collected the data, and we were both heavily involved in the data analyses and interpretation.

4.1 Introduction

Autobiographical memories form the narrative of our lives. These episodic memories of personal past experiences are known to depend on the hippocampus during initial encoding (Scoville and Milner, 1957; Spiers et al., 2001; Cipolotti and Bird, 2006), but their subsequent neural fate is less certain. Understanding how remote autobiographical memories are supported by the brain, and how it is we can vividly re-experience episodes from decades earlier are key questions at the heart of memory neuroscience. Consolidation of memories over time undoubtedly occurs at the synaptic level (Dudai, 2004). By contrast, systems-level consolidation (Dudai, 2004), and whether memories gradually lose their dependence on the hippocampus in favour of support by neocortical regions, is still vigorously debated.

As outlined in Chapter 1, several theoretical accounts dominate the literature on systems-level consolidation. The standard model of consolidation (SMC) has the greatest longevity (Scoville and Milner, 1957; Marr, 1971; Squire and Alvarez, 1995; Squire et al., 2004), and proposes that when an episodic (including autobiographical) memory is first formed, it is represented in the neocortex, but depends on the hippocampus for retrieval. Over time connections between the cortical regions that support the memory strengthen. As this happens the memory becomes less dependent on the hippocampus until eventually it is fully consolidated, no longer requiring hippocampal involvement during retrieval. Various alternative theories argue that the hippocampus is always necessary for the support of truly vivid autobiographical memories regardless of remoteness (Nadel and

Moscovitch, 1997; Moscovitch et al., 2005; Hassabis and Maguire, 2007, 2009; Winocur and Moscovitch, 2011), although they do not rule out a degree of cortical consolidation over time in some circumstances. These theories, including multiple trace theory (Nadel and Moscovitch, 1997; Moscovitch et al., 2005), its more recent incarnation the transformation hypothesis (Winocur et al., 2010; Winocur and Moscovitch, 2011), and also the scene construction theory (Hassabis and Maguire, 2007, 2009) were born out of problems noted with the SMC, not least of which related to retrograde amnesia.

That some patients with lesions apparently restricted to the hippocampus show limited retrograde amnesia, is the cornerstone of SMC. The patients' impairment in recalling recent but not remote memories is held to demonstrate the time-limited role of the hippocampus (e.g. Zola-Morgan and Squire, 1990; Bayley et al., 2005). However, some patients also with lesions purportedly restricted to the hippocampus do not show a graded retrograde amnesia, but instead are amnesic for both recent and remote autobiographical memories (for a full review see Winocur and Moscovitch, 2011). Moreover, even where a graded RA occurs, the length of RA (sometimes decades) can be difficult to understand mechanistically (Nadel and Moscovitch, 1997; Moscovitch et al., 2005). Differences in the memory tasks used, the scoring methods applied, the nature and extent of patients' lesions, and the appropriateness of the control participants, continue to confound attempts to elucidate the neural basis of remote autobiographical memories. Studying the retrieval of recent and remote autobiographical memories in healthy participants using standard fMRI circumvents some of

these issues, but findings in this domain are not clear-cut either, with some studies reporting hippocampal activations for remote memories (e.g. Maguire et al., 2001; Ryan et al., 2001; Maguire and Frith, 2003; Piolino et al., 2004; Rekkas and Constable, 2005; Steinvorth et al., 2006; Viard et al., 2007) and others not (Niki and Luo, 2002; Maguire and Frith, 2003; Piefke et al., 2003).

The debate, therefore, rumbles on with apparently no resolution in sight. Here we attempted to break the deadlock by employing MVPA to examine the question of memory remoteness in a different way. In the previous chapter, I showed that it is possible to predict or 'decode' episodic-like memories of short movie clips viewed prior to scanning from patterns of activity across voxels in the hippocampus. To date, MVPA has not been applied to autobiographical memory and yet the ability to examine the representation of information relating to specific stimuli (e.g. memories) in particular brain areas afforded by MVPA makes it an especially suitable approach for investigating autobiographical memory.

We therefore used high resolution fMRI (Carr et al., 2010) and MVPA to examine whether information pertaining to recent and remote autobiographical memories was present in the hippocampus and other areas of interest. We achieved high resolution scanning by acquiring images in a limited volume that included the temporal lobes (medial and lateral), and also retrosplenial cortex and ventro-medial prefrontal cortex (see Chapter 2). The latter was important to include, because this area has been implicated in memory consolidation (Nieuwenhuis and Takashima, 2011). In order to

make a clear distinction between recent and remote memories, the recent memories we examined were two weeks old, and the remote memories were ten years old. We carefully matched the recent and remote memories on variables such as frequency of recall, vividness, level of detail, and emotional valence, in order to rule these out as explanatory factors in our analyses.

We reasoned that if the SMC was correct (Squire, 1992; Squire et al., 2004), and the hippocampus has no role to play in supporting remote autobiographical memories, then there would be no information relating to such memories present there – why would there be if the hippocampus is no longer required? In this case, the classifier performance should be at chance for the remote memories, while in contrast, information relating to the fresh recent memories would still be present in the hippocampus and it should be possible to decode these recent memories from patterns of activity across voxels in the hippocampus. Moreover, it should be possible to decode the remote memories from patterns of activity in cortical areas, given the SMC's position that remote memories are consolidated there. On the other hand, the prediction of the alternative theories (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Hassabis and Maguire, 2007, 2009; Winocur and Moscovitch, 2011) would be that decoding of recent *and* remote autobiographical memories should be possible from patterns of activity across voxels in the hippocampus, given their view that rich and vivid memories of this type always depend on the hippocampus regardless of remoteness.

4.2 Methods

4.2.1 Participants

Twelve healthy right-handed, university-educated, participants (9 female) took part in the experiment (mean age 27.5 years, SD 3.2, range 22-33). All had normal or corrected-to-normal vision and gave informed written consent to participation in accordance with the local research ethics committee.

4.2.2 Autobiographical Memories

One week before scanning twelve participants were interviewed and asked to recall specific events that could be re-experienced vividly from two time periods, two weeks ago (recent) and 10 years ago (remote). The recent memories were on average 13.3 (SD 2.7) days old, while the remote memories were on average 10.4 (SD 0.57) years old. Participants also rated each memory along a range of dimensions which are detailed in Table 4 (in section 4.3). Three memories from each time period were selected and matched using these ratings, resulting in 6 memories for use during scanning.

4.2.3 Pre-scan training

On the day of scanning participants were trained to recall each memory within a 12 second recall period after viewing a word cue. There were six training trials per memory. They were encouraged to make each recall as vivid as possible and to maintain the consistency of recall of each memory for the duration of the 12 seconds.

4.2.4 Task

During scanning participants recalled each memory fourteen times. On each trial, a verbal cue was presented which indicated which of the six memories the participant was required to recall (see Figure 19). Following this, an instruction appeared on the screen indicating that the participant should close their eyes and vividly recall the cued memory. Participants were instructed not to begin the recall process until this instruction appeared, and were trained on this procedure in the pre-scan session. After 12 seconds, an auditory tone sounded signalling they should open their eyes. After this, the participant was required to provide ratings about the preceding recall trial using the five-key button-box. Firstly, they rated how vivid the preceding recall trial was (scale: 1 – 5, where 1 was not vivid at all, and 5 was extremely vivid). Secondly, they rated how consistently they had recalled it relative to the original event (scale: 1 – 5, where 1 was not consistent at all, and 5 was extremely consistent). These ratings were used to select only the most vivid (ratings of 4 or 5) and most consistently recalled (ratings of 4 or 5) for inclusion in the MVPA analyses. There were a total of 84 trials, with fourteen trials of each memory presented in a pseudo-random order, whilst ensuring that the same memory was not repeated twice or more in a row. In order to assess the degree to which the memories had been retrieved since the pre-scan session, after scanning each participant was asked: “During the scan did you think about the interview last week?”, where 1 was not at all…5 all the time. They were also asked “Do you feel that repeatedly recalling a memory changed the memory in any way?”, where 1 was not at all…5 very much.

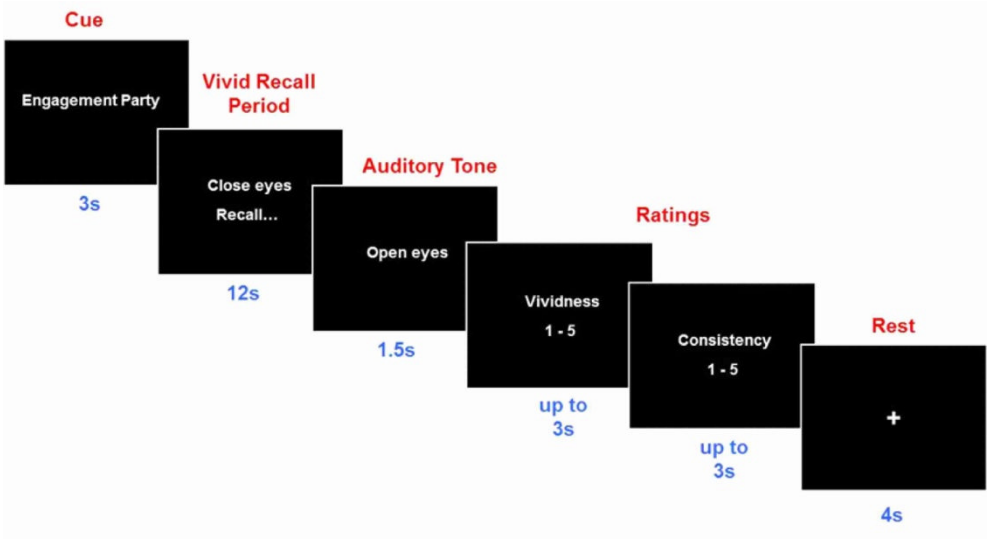

Figure 19. Example timeline from a trial during scanning. *On each trial participants saw a cue telling them which memory to recall. They then closed their eyes and proceeded to recall the memory as vividly as possible. After 12 seconds an auditory tone sounded signalling they should open their eyes, and they then made ratings of how vividly the memory had been recalled and also how consistently they had recalled it relative to the original event.*

4.2.5 Image acquisition

All functional images were acquired using the high-resolution fMRI sequence, taking a partial volume through the MTL (see Figure 20). Field maps were acquired for distortion correct. See Chapter 2 (Methods) for details of each of these scanning sequences. In addition to the functional scans, two structural images were acquired. The first was a whole brain T1-weighted 3D FLASH sequence (resolution 1 x 1 x 1 mm) which was acquired immediately following the functional scan. The second was the T2-weighted, high-resolution, sub-millimetre sequence described in Chapter 2 (Methods), and this was acquired in a separate scanning session. To improve the SNR of this anatomical image, four scans were acquired for each participant, coregistered and averaged.

Field inhomogeneities in the human brain can result in local signal loss and reduction in BOLD sensitivity which can be compensated by use of z-shim gradients (Deichmann et al., 2003; Weiskopf et al., 2006). However, the choice of an optimal z-shim value can be challenging when several brain regions with different field inhomogeneities are involved. In this experiment the functional imaging was slightly modified in order to try and reduce signal loss in the various regions of interest (see section 4.2.6 and Figure 20). In order to do this, we assigned an optimal z-shim value to each slice of the encoding volume; accounting for all the regions involved this study. The resulting set of optimal z-shim values was used in all subsequent fMRI runs. In order to calculate the optimal z-shim values, a test scan was acquired for each participant before the fMRI experiment. For this scan, an EPI volume was acquired with z-shim values ranging from -5mT/m*ms to 4mT/m*ms in steps of 0.2 mT/m*ms. All other acquisition parameters were kept identical for the fMRI acquisitions. Regions of interest (ROIs) were manually defined for each participant. For each slice of the EPI volume, the signal averaged over all the voxels present in the ROIs was calculated and the optimal z-shim value yielding maximum signal was selected. For slices that did not contain any ROI, the optimal z-shim value was set to zero. A Butterworth low pass filter was used (cut off frequency of 0.3) to smooth the distributions of optimal z-shim values in order to avoid large changes in signal between neighbouring slices due to sudden changes in optimal z-shim values. Before the main scanning experiment, a baseline session comprising 100 volumes without z-shim manipulation was undertaken. We used this baseline to measure the BOLD signal change when z-shim manipulation was utilized. A signal increase of between 1% and 4% was noted over all

regions. A significant signal increase in temporal poles of 18.25% (SD 10.22) was also observed. Therefore, the z-shim manipulation allowed us to obtain a significant signal increase in the anterior temporal lobes without any signal loss in other regions of interest.

4.2.6 ROI segmentation

Manual segmentation of relevant brain regions (see Figure 20) was performed using ITK-SNAP (Yushkevich et al., 2006) on the T2 high-resolution structural images. Hippocampal anatomy (HC) was identified using the Duvernoy hippocampus atlas (Duvernoy, 2005). The entorhinal/perirhinal cortex (EPC), parahippocampal cortex (PHC) and temporal pole (TP) were segmented according to the protocol described in Insausti et al. (1998). The lateral temporal cortex anatomy (LTC) was identified using the Duvernoy whole brain atlas (Duvernoy, 1999), and the retrosplenial cortex (RSC) was identified as BA region 29 and 30 (Vann et al., 2009). Ventro-medial prefrontal cortex (vmPFC) segmentation was identified as the region where previous work has suggested increased involvement in for consolidation (Nieuwenhuis and Takashima, 2011), namely BA 25, ventral parts of areas 24 and 32, the caudal part of area 10, and the medial part of area 11. Intra-rater reliability was calculated using the DICE overlap metric, defined as the volume of overlap between two regions of interest, divided by the mean volume (Dice, 1945). This produces an overlap measure between 0 and 1, where 0 signifies no overlap and 1 is a perfect match. Heidi Bonnici performed intra-rater reliability with a 6-month interval between first and second segmentations. The DICE metric results were: HC 0.90, EPC 0.77, PHC 0.82, RSC 0.70, TP 0.85, LTC 0.77,

and vmPFC 0.78.

For segmentation of the hippocampus into its anterior and posterior parts, we based our segmentation protocol on that of Hackert et al. (2002), where the anterior 35% of the hippocampus was labelled as anterior and the remainder as posterior.

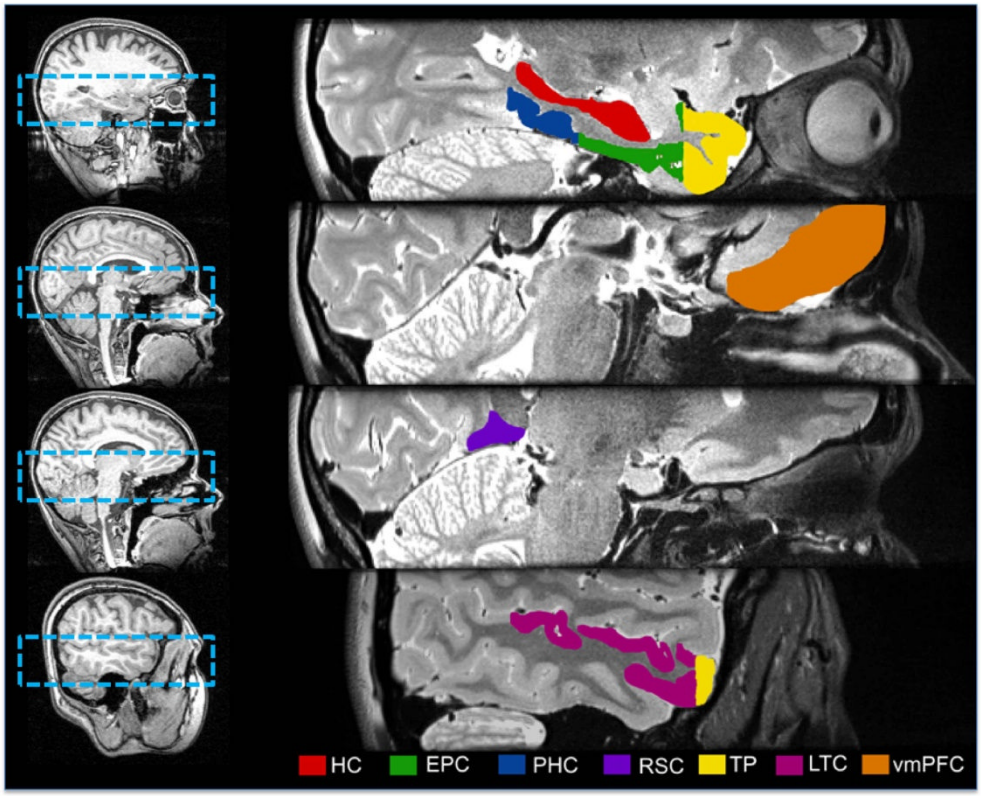

Figure 20. The brain regions examined. *High resolution functional (1.5mm isotropic voxels) and structural (0.5mm isotropic voxels; right column) MRI scans were acquired in a limited volume (see left column). The following regions were delineated bilaterally: hippocampus (HC), entorhinal and perirhinal cortices (EPC – data relating to these two regions were amalgamated as they showed very similar profiles), parahippocampal cortex (PHC), retrosplenial cortex (RSC - BA 29,30), temporal pole (TP), lateral temporal cortex (LTC - middle temporal gyrus), and ventro-medial prefrontal cortex (vmPFC – including BA 25, ventral parts of areas 24 and 32, the caudal part of area 10, and the medial part of area 11).*

4.2.7 Image pre-processing for MVPA analysis

The first six EPI volumes were discarded to allow for T1 equilibration effects (Frackowiak et al., 2004). The remaining EPI images were then realigned to correct for motion effects, and minimally smoothed with a 3mm FWHM Gaussian kernel. This minimal smoothing was included in order to reduce noise from potential residual misalignments between scans, while still ensuring that information was present at a fine-grained spatial resolution. A linear detrend was run on the images to remove any noise due to scanner drift (LaConte et al., 2005). Next the data were convolved with the canonical haemodynamic response function (HRF) to increase the signal-to-noise ratio (Frackowiak et al., 2004). This HRF convolution effectively doubled the natural BOLD signal delay, giving a total delay of approximately 12s. To compensate for this delay, all onset times were shifted forward in time by three volumes, yielding the best approximation to the 12s delay given a TR of 3.5s and rounding to the nearest volume (Haynes and Rees, 2006). Functional volumes were extracted from 12 second period of vivid recall of each trial (Figure 19).

4.2.8 MVPA classification

The MVPA classification analysis was identical to that described in the previous chapter, with the one modification that a 10-fold cross-validation scheme was used rather than leave-one-out (see Chapter 2 and Duda et al., 2001). This change was included in order to reduce processing time (which can be substantial using complex MVPA approaches such as this).

MVPA decoding was applied separately in the two memory conditions (recent and remote). Thus, in each case, the classifier was decoding between three individual memories, and we ended up with a single accuracy value for each condition and each ROI. These accuracy values represent the amount of distinct information that was present about the three individual memories from each condition separately. This allowed us to ask two important questions. First, for each condition separately, is there any evidence for information about the individual memories in each of the ROIs. Second, by comparing the accuracies between conditions, we can determine whether the strength of episodic representation changes as a function of time.

4.2.9 Information maps

Information maps were generated from the feature selection results of each ROI, as described in the previous chapter. Separate information maps were generated for each of the memory conditions (recent and remote), and these were then superimposed on 3D images of participants' hippocampi in order to view the locations of the voxels containing the most relevant information for the two types of memory.

To measure the overlap between recent and remote memory information maps for each participant we used the DICE metric (Dice, 1945). To test any overlap against chance, we randomly shuffled the positions of the recent and remote maps within the hippocampus 1000 times, and every time measured the overlap. This provided us with a null distribution of the DICE metric. We could then test the actual overlap directly against this null distribution using a one-way t-test.

4.2.10 Statistical analysis

The values from each brain region were compared to chance using t-tests. Within each brain region, the values for recent and remote memories were examined using repeated measures ANOVAs. Brain regions were not directly compared to each other in this study due to potential problems with interpretation given the differences in their size (see section 3.4 in Chapter 3, and Diedrichsen et al., 2011). A threshold of $p<0.05$ was employed throughout.

4.3 Results

4.3.1 Behavioural Results

Table 4 displays the ratings for each memory type. Notably, we found no significant difference between the two types of memory in any of the variables, demonstrating that these extraneous factors were controlled across the recent and remote memories. Importantly, the participants hardly thought about the memories at all between the interview and the scan. Additionally, the memories were rated as not having been recalled very much since the initial occurrence of the event, suggesting that none of these memories were over-rehearsed "semanticized" memories. After scanning, participants were asked: "During the scan did you think about the interview last week?", where 1 was not at all…5 all the time. Participants did not think about the interview during scanning [1.08 (0.29)]. They were also asked "Do you feel that repeatedly recalling a memory changed the memory in any way?", where 1 was not at all…5 very much, and indicated that the

memories were hardly changed by multiple repetitions [2.08 (0.79)]. When trials that were not vivid or consistent were excluded (see Methods), this resulted in 11.58 (0.30) trials on average for each recent memory and 10.14 (0.89) trials on average for each remote memory being entered into the MVPA analysis.

Variable	Recent mean (SD)	Remote mean (SD)	Recent vs Remote t value	Recent vs Remote p value
Recall frequency before the interview	1.64 (0.611)	1.83 (0.415)	1.258	0.235
Recall frequency between the interview and scan	1.08 (0.208)	1.03 (0.095)	1.483	0.166
Vividness	4.58 (0.352)	4.39 (0.372)	1.549	0.15
Level of detail	4.47 (0.414)	4.14 (0.576)	1.7	0.117
1st/3rd person perspective	1 (0)	1.08 (0.149)	1.915	0.082
Emotional valence	3.17 (0.301)	3.14 (0.172)	0.372	0.717
Active/static event	1 (0)	1.03 (0.095)	1	0.339
Consistency of recall trial-to-trial	4.83 (0.225)	4.72 (0.372)	1.317	0.215

Table 4. Memory characteristics. *Ratings were on a scale of 1 to 5, where 1 was the minimum and 5 the maximum. For emotionality: 1,2 = negative, 3 = neutral, 4,5 = positive. For 1st/3rd perspective: 1 = 1st person, 2 = 3rd person. For active/static event: 1 = active, 2 = static.*

4.3.2 MVPA analysis

The results for the left and the right hemispheres were highly similar, and so all analyses reported here are based on the average decoding accuracies across hemispheres. We first explored whether it was possible to predict which of the recent memories was being recalled solely from the pattern of activity across voxels. MVPA classifiers operating on voxels in all seven regions were able to distinguish between the three recent autobiographical memories significantly above chance (see Figure 21, blue line):

hippocampus (HC): t=3.463, p=0.005; entorhinal/perirhinal cortex (EPC): t=3.431, p=0.006; parahippocampal cortex (PHC): t=3.209, p=0.008; retrosplenial cortex (RSC): t=7.639, p=0.001; temporal pole (TP): t=3.499, p=0.005; lateral temporal cortex (LTC): t=4.19, p=0.002; and vmPFC: t=3.35, p=0.006. This initial result therefore extends the results from my previous chapter, and demonstrates that this method can be applied to genuine autobiographical memories as well as more controlled episodic representations.

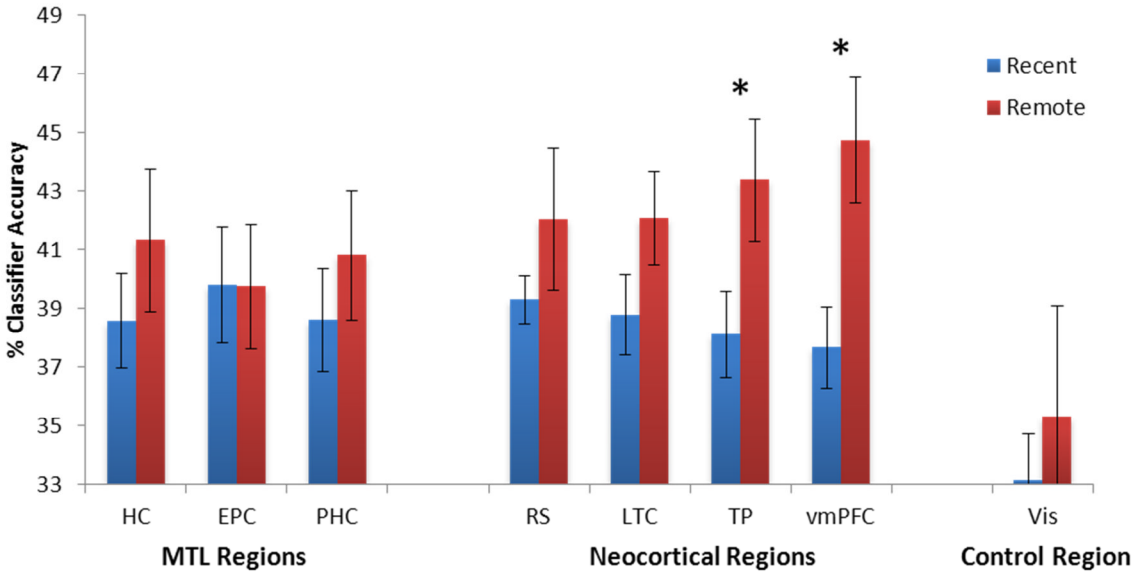

Figure 21. MVPA results for recent and remote autobiographical memories. *Hippocampus (HC), entorhinal and perirhinal cortices (EPC), parahippocampal cortex (PHC), retrosplenial cortex (RSC), temporal pole (TP), lateral temporal cortex (LTC), and ventro-medial prefrontal cortex (vmPFC) were examined. Medial temporal regions, including the hippocampus, contained similar amounts of information about recent and remote autobiographical memories, while cortical areas (other than retrosplenial cortex) contained more information relating to remote memories. *P<0.05; chance is 33%. The difference between recent and remote memories just failed to reach statistical significance for LTC. Error bars represent +/- 1 standard error of the mean.*

Having established that predictable information was present in our regions of interests, including the hippocampus, that enabled above-chance decoding of the recent autobiographical memories, we next considered the three remote memories. Again, MVPA classifiers operating on voxels in all seven regions were able to distinguish between the three remote autobiographical memories significantly above chance (see Figure 21, red line: HC: t=3.426, p=0.006; EPC: t=3.175, p=0.009; PHC: t=3.548, p=0.005; RSC: t=3.713, p=0.003; TP: t=4.966, p=0.001; LTC: t=5.669, p=0.001; and vmPFC: t=5.49, p=0.001). Our results, therefore, show that information about the remote memories was represented not only in cortical areas, but also in the medial temporal lobe, including the hippocampus.

We next used F-tests to test for a significant difference in classification accuracy for the recent and remote memories. We performed this analysis separately on the three MTL regions (HC, EPC, PHC) and the four cortical areas (RSC, TP, LTC, vmPFC) in order to separately assess the effect of memory remoteness on these two regions. This analysis revealed no significant effect of memory remoteness in the MTL ($F = 0.40$, $p=0.54$), while the cortex showed a clear significant effect, with higher decoding accuracies in the remote memory condition ($F = 6.79$, $p=0.038$). Post-hoc analyses revealed that the difference between recent and remote memory decoding was most apparent in the TP (t=-2.029, p=0.033) and vmPFC (t=-2.833, p = 0.008). A similar trend was observed in LTC (t=-1.457, p=0.087), and no difference between recent and remote memories in RSC (t=-1.179; p=0.132). Put together, these results suggest a pattern whereby MTL regions contain similar amounts of information about recent and remote

autobiographical memories, while cortical regions (except RSC) show an increase in episodic information over time.

We specifically focused on brain regions within our partial volume that are known to be involved in autobiographical memory retrieval. However, we also examined accuracy values in control (i.e. not memory-related) cortical regions (left and right lateral posterior visual cortex). Classifier accuracies for recent and remote memories were at chance, i.e. it was impossible to predict which memories were being recalled from the patterns of activity across voxels there (collapsed across left and right posterior visual cortex; recent memories p=0.9; remote memories p=0.6). This shows that our classification analysis was not biased toward findings above-chance accuracies.

4.3.3 Spatial distribution of information within the hippocampus

Given our particular interest in the hippocampus, and our finding that information relating to both recent and remote memories was represented there, we next considered whether the voxel patterns (and by inference the underlying neuronal populations) that supported the recent memories overlapped with those supporting the remote memories in the hippocampus. Information maps for recent and for remote memories were created that comprised the voxel sets that produced above-chance classification accuracy (see Methods and Figure 22), and the overlap between these information maps was measured using the DICE metric. The overlap for the recent and remote memory information maps in the hippocampus was strikingly low

(0.18). Indeed, when tested against the permuted null distribution (see Methods), this overlap was significantly lower than would be expected by chance (t=-3.216, p=0.004). Notably, when we repeated this analysis in cortical areas, none produced an overlap measure different to chance (TP: t=-0.714, p=0.245; LTC: t=1.089, p=0.150; vmPFC: t=-0.554, p=0.295).

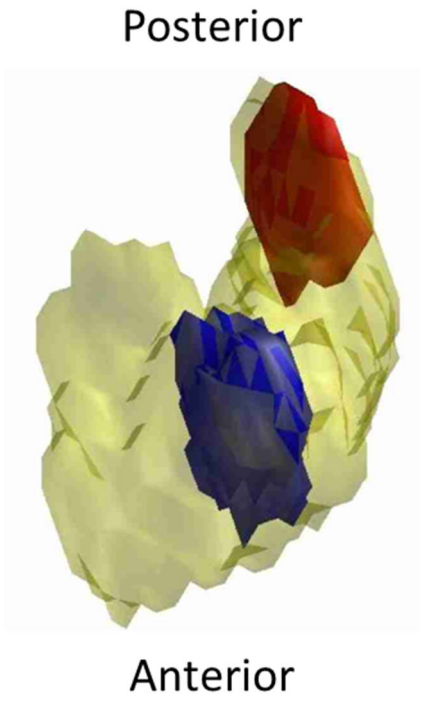

Figure 22. Information maps in the hippocampus*. Information maps for recent and remote autobiographical memories were created comprising the voxel sets that produced above-chance classification accuracy (see Methods). An example 3D rendered left hippocampus chosen at random from one of the participants is shown: Red=remote memories, blue=recent memories. The lack of overlap between the information maps is clearly apparent.*

This suggests that there is physical separation of neuronal populations involved in representing recent and remote autobiographical memories within the hippocampus. Thus, while involved in the representation of both types of memory, the location of the representations within the hippocampus appears to be distinct. If this is the case, then there could be a bias towards different regions of the hippocampus for these two types of memory. Visual

inspection of the information maps of the participants (e.g. Figure 22) suggested a bias towards the anterior hippocampus for recent memories, and the posterior for remote memories. To investigate this formally, the hippocampus was subdivided into anterior and posterior portions, and the MVPA decoding analysis for both recent and remote memories was repeated in each hippocampal subdivision (Figure 23). Above-chance classification was apparent for anterior and posterior hippocampus for recent and remote memories, showing that information about both types of memory was represented in both portions of the hippocampus (recent memories: anterior: t=2.561, p=0.026; posterior t=2.242, p=0.047; remote memories: anterior: t=4.665, p=0.001; posterior: t=4.225, p=0.001).

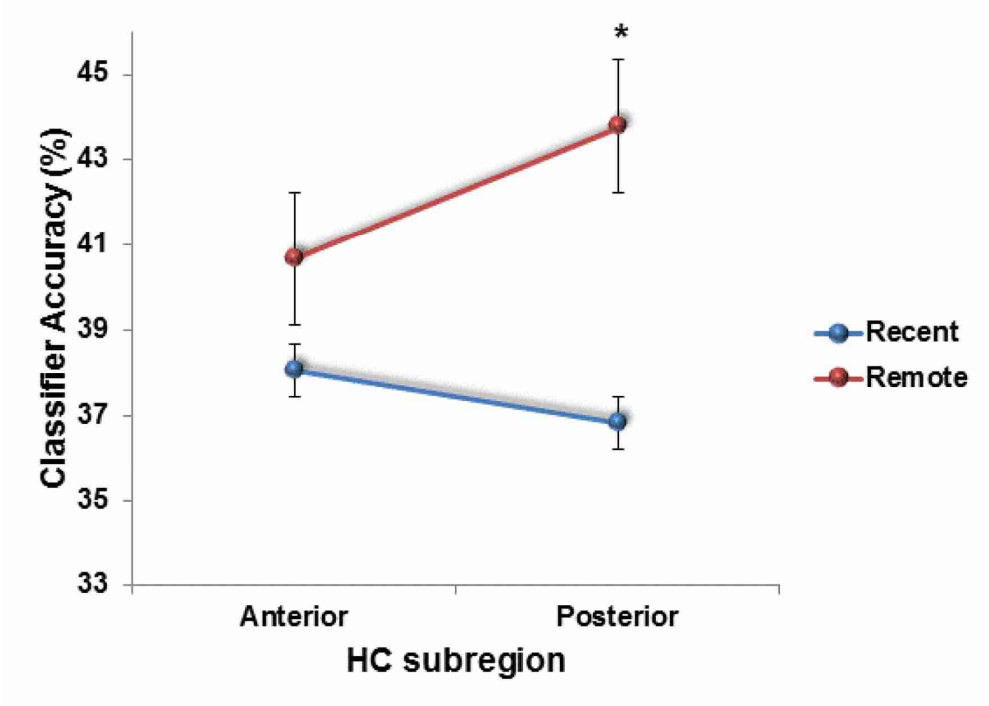

Figure 23. MVPA results for anterior and posterior subregions of the hippocampus. *The posterior hippocampus contained more information relating to remote memories than recent, while in anterior hippocampus, there was no significant difference in the amount of information for the two types of memory. *P<0.05; chance is 33%. Error bars represent +/- 1 standard error of the mean.*

The key question was whether a systematic bias towards one or other type of memory within the sub-divisions existed that would result in a difference in classification performance. This is indeed what we found, with the posterior hippocampus containing more information relating to remote memories compared to recent ($t=-2.852$ $p=0.008$), while in anterior hippocampus, there was no significant difference in the amount of information for the two types of memory ($t=-0.986$, $p=0.173$).

4.4 Discussion

In this study we used MVPA and high resolution fMRI to examine whether information pertaining to specific recent and remote autobiographical memories was present in the hippocampus, adjacent medial temporal lobe (MTL) structures, and cortical areas within temporal and frontal regions. There were three main findings. First, discernible and predictable information about both recent and remote autobiographical memories was present in the hippocampus, with no difference between the classification accuracies for the two types of memory. Second, cortical regions also contained information about both recent and remote autobiographical memories, but the classification accuracy was significantly higher for remote memories, particularly in the temporal pole and vmPFC. Third, even though the hippocampus contained information about recent and remote autobiographical memories in apparently equal measure, the information had a spatial bias, with significantly higher classification accuracy in the posterior hippocampus for remote compared to recent memories.

Before discussing the theoretical implications of these findings, it is important to consider whether factors other than the recency/remoteness of autobiographical memories could have influenced our results. For instance, during scanning perhaps participants were recalling the pre-scan interview where the memories were initially elicited. However, when they were asked post-scan '*During the scan did you think about the interview last week*?' the over-whelming response was 'not at all'. Moreover, the interview concerned both the recent and remote memories, and thus this common interview experience cannot explain the differential decoding effects we found within posterior hippocampus and the cortex. Another common criticism of fMRI studies of remote memory concerns the possibility of re-encoding activity (Squire et al., 2004). The argument here is that when participants are retrieving remote memories in the scanner, they will also be encoding the experience of retrieving these memories. In other words, any activity found within the hippocampus during this condition may be due to encoding new memories rather than retrieving old memories. Importantly this argument does not hold for the current results due to the fact that the retrieval of each individual remote memory elicited the same pattern of activation across multiple retrieval trials (this is necessarily the case, or else MVPA classification would have been at chance level). If the participants were encoding new memories on each trial, then we would expect to see distinct patterns of activation on every trial, and MVPA classification would have failed. It could also be the case that recalling a remote memory re-activated it, effectively transforming it back into a recent memory. If this was the case, then the prediction would be of no difference between recent and remote memories (if all memories were now essentially recent). However,

the differential effects, cortically and within the hippocampus itself, for recent and remote memories underscore the distinction between the memory types, and render this explanation unlikely. Thus, none of these alternatives can adequately explain this set of results, giving us confidence that the effects directly reflect episodic information related to the retrieval of recent and remote memories.

Our results have direct relevance for the debate about systems-level consolidation of memory representations. The presence of information encoded in the pattern of activity in the hippocampus concerning both recent and remote rich and vivid autobiographical memories speaks against the SMC (Squire, 1992; Squire et al., 2004) but in favour of alternative accounts such as MTT (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Winocur and Moscovitch, 2011), and scene construction theory (Hassabis and Maguire, 2007, 2009). On the other hand, the increase in classification accuracy in neocortical regions during recall of remote memories is in line with the predictions of SMC, while MTT also allows for cortical consolidation to occur. No theoretical position predicted the intra-hippocampal distinction for recent and remote memories that we uncovered. The SMC asserts the hippocampus is out of the loop for retrieving remote memories, and the alternative views, while believing the hippocampus to be involved in perpetuity, are not specific with regards to what might occur within the hippocampus for the two types of memory. Our findings, therefore, constrain and broaden the current view of systems-level consolidation. We provide strong evidence that recall of rich and vivid autobiographical memories involves the hippocampus regardless of

remoteness. Nevertheless, changes occur in the neocortex, such that remote memories are more strongly represented there. We also now show that the posterior hippocampus, just like the neocortex, respects the distinction between recent and remote autobiographical memories.

This clearly begs the question, what exactly differs between the representations of recent and remote autobiographical memories? The memories studied here were selected carefully to be matched on a range of variables (Table 4), and were vividly recalled and re-experienced. Phenomenologically, therefore, no obvious differences were apparent between recent and remote memories that can easily explain the neural distinctions we observed. It has been proposed that remote memories can become semanticized over time (Winocur and Moscovitch, 2011), transforming into more gist-like versions that are represented in extra-hippocampal structures. This appears to be at odds with our findings, which show increased information in neocortical regions for remote autobiographical memories that retained their vividness, could be richly re-experienced, were not gist-like in nature, and that were still also represented in the hippocampus. Unless the remote memories in our study have been semanticized in some way that did not affect their phenomenological qualities, which seems unlikely, then semanticization cannot explain our findings. However, Winocur and Moscovitch (2011) claim that detailed and gist-like representations of the same memory can co-exist. Perhaps the gist-like version is responsible for the increased information in neocortical areas that we observed, although the idea of maintaining two forms of the same memory (when one of these is the fully detailed and vivid form) seems

somewhat odd in terms of neural efficiency. Moreover, this account does not clearly explain the posterior hippocampal bias for remote compared to recent autobiographical memories.

Remote memory representations may be more streamlined and neural coding more efficient than recent memories, and perhaps this made it easier for the classifier to detect memory-specific information. While this may be the case in cortical regions such as vmPFC, if streamlining improved information detectability then increased classification accuracy for remote memories should have been apparent across the hippocampus. Thus, it is not clear why more streamlined memories would be represented to a greater degree in one particular part of the hippocampus. Related to this is the idea that remote memories are largely consolidated to the neocortex and what remains in the hippocampus is a distilled version or index that identifies a memory and is somehow involved in retrieval (Teyler and DiScenna, 1985). Our data would then suggest that these indices for remote memories are located preferentially in the posterior hippocampus, although it is not obvious why this would be the chosen site. However, the opposite can also be argued, that remote memories, rather than being streamlined, actually have more associations and are embedded into existing schema (Tse et al., 2007, 2011; van Kesteren et al., 2010; McKenzie and Eichenbaum, 2011) to a greater degree than recent memories. This could explain the increased information detected by the classifier in cortical areas and the hippocampus although, as with the streamlining idea above, it is not clear why the increased information would be located in the posterior hippocampus in particular.

By drawing on extant theories and our data, and considering the brain areas in question and what is known about their functions, we would like to suggest another possible explanation for our findings, and for system-level consolidation. In neocortical areas, classification accuracies were significantly above chance even for the recent memories, showing that quite soon after the event occurred, cortical representations of some form were already established (Tse et al., 2007, 2011; Sharon et al., 2011). The increased classification accuracies in these neocortical areas for remote memories, suggests that more information about these memories is represented there, in keeping with a view that memory details and content have been transferred to these regions and reside there (Squire, 1992; Squire et al., 2004). The temporal poles and also the lateral temporal cortex (which just failed to reach significance for remote memories) are known to be storage sites for semantic information, as evidenced most clearly by the loss of such information in semantic dementia (Hodges et al., 1992). Patterns of activity across voxels in the vmPFC led to the highest decoding accuracies. This region has been linked to memory consolidation in a number of other studies [e.g. Bontempi et al. (1999); Tse et al. (2011); see Nieuwenhuis and Takashima (2011) for a full review], although typically it has been cast as the 'new controller' when the hippocampus is no longer part of the processing loop. However, our data suggest that the hippocampus retains involvement in remote memories. We therefore propose that at least one function of vmPFC may also be as a storage site for content and details of remote autobiographical memories.

We speculate that what specifically happens is this: very recent memories are experienced largely as coherent scenes/events that are temporarily represented in the hippocampus (utilising anterior and posterior aspects and their respective functions), with transfer to and consolidation within the neocortex happening quickly and from an early stage. The constituent elements of the autobiographical memories are then the preserve of the neocortex. At retrieval, and by default, this piecemeal information is automatically funnelled back into the hippocampus, but in order to be assembled into a coherent form, this requires particular input from a process that is performed in the posterior hippocampus. This, we suggest, is why the remote memories were discernible to a greater degree in posterior hippocampus, because they rely on this process more than recent memories. For some memories, the information that the hippocampus receives will lack contextual or other details and it cannot be reconstructed to the point of being vividly re-experienced. These memories will remain gist-like or semantic. Functional differentiation down the long axis of the hippocampus has been well documented (Moser and Moser, 1998; Maguire et al., 2000; Gilboa et al., 2004; Kahn et al., 2008; Fanselow and Dong, 2010; Poppenk and Moscovitch, 2011). In particular, the posterior hippocampus has been associated with spatial processing (e.g. Moser and Moser, 1998; Maguire et al., 2000). We speculate that the posterior hippocampus may facilitate the spatial framework into which the elements of a remote memory are bound and re-constructed (Hassabis and Maguire, 2007, 2009), in line with findings from patients with hippocampal damage who have lost the ability to construct spatially coherent scenes (e.g. Hassabis et al., 2007; Race et al., 2011).

This account clearly provokes further questions. In particular, how do relevant instead of random elements get channelled back into the hippocampus during retrieval? The elements of a particular memory likely remained linked in a low-level manner when stored neocortically, and this may be mediated by vmPFC (Nieuwenhuis and Takashima, 2011). Perhaps it (or sub-regions within it) has a dual role that involves both storing memory components (see above), but also suppressing those that are not relevant in order to convey only the pertinent set of information back to the hippocampus (Goshen et al., 2011; Nieuwenhuis and Takashima, 2011). Patients with damage involving this region can suffer from confabulation (Schnider, 2003) perhaps because the ability to suppress irrelevant information is lost. That a brain region might play more than one role is often overlooked when theorising about memory. Our suggestion here that the hippocampus acts as an encoder and temporary memory processor on the one hand, and also as a spatially-based reconstruction device (Hassabis and Maguire, 2007, 2009) may serve to explain some of the discrepancies in findings across patients with hippocampal damage and amnesia, depending on the extent and location of hippocampal damage (Martin et al., 2011; Winocur and Moscovitch, 2011). If damage to the hippocampus leaves enough of the posterior portion intact, remote memories could be spared (Squire, 1992), while more complete hippocampal damage would impair retrieval of both recent and remote autobiographical memories (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Hassabis and Maguire, 2007, 2009; Winocur and Moscovitch, 2011).

The issue of systems-level consolidation is at the core of memory neuroscience but has so far eluded agreement. By adopting a different approach using MVPA and high resolution fMRI we were able to offer a new perspective on the representation of individual recent and remote autobiographical memories in medial temporal and cortical areas. Here we focused on vivid and easily-retrievable memories, and at two disparate timescales. In the future it will be necessary to examine memories that vary in their vividness and age in order to get a more complete picture of the system at work. It will also be essential to consider the role of individual hippocampal subfields in supporting recent and remote autobiographical memories, as currently there is almost a complete absence of such information in humans (Bartsch et al., 2011; Goshen et al., 2011). Finally, future work will need to explore the processes that underpin the increase in the strength of remote memory representations in the cortex. Similarly, it will be important to elucidate precisely why the posterior hippocampus shows an increase in representational strength with remote memories. The deployment of MVPA analyses in combination with well-controlled experimental design (e.g. longitudinal MVPA studies investigating the representation of specific memories over time) may help to shed light on these issues.

5 Chapter 5

Decoding overlapping memories

Precis

The previous two experiments demonstrated that it is possible to decode individual episodic and autobiographical memories from patterns of activity across voxels in the hippocampus. In each of these studies, the memories differed along a variety of different dimensions including spatial location, their content and the people involved. It is therefore possible that the MVPA analyses could be detecting any one of these sources of information (or a combination of them) in order to decode the memories. Thus, it was not possible to determine exactly what information was being decoded in these previous studies, which limits our ability to draw inferences about the nature of the episodic representations themselves. In the next experiment, I developed a new paradigm that permitted more control over the constituent elements of each episode, thereby allowing me to determine more precisely the nature of the hippocampal representations.

5.1 Introduction

Our daily lives usually involve encounters with a relatively limited range of people and locations, and consequently, the episodic memories that are formed often contain much overlap. Nevertheless, most of the time, we are able to remember each event as a distinct episode. The hippocampus has long been implicated as the critical brain structure involved in separating overlapping episodes into unique representations, which are then stored as distinct memory traces (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Whilst the theoretical basis for

this idea has a strong grounding in the anatomy of the hippocampus and in the rodent literature (Lee et al., 2004; Leutgeb et al., 2004, 2007; Vazdarjanova and Guzowski, 2004), empirical evidence for the existence of traces of complex episodic memories in the human hippocampus remains scarce.

Recent studies (see Chapters 3 and 4) demonstrated that specific episodic-like memories can be decoded solely from patterns of fMRI activity across voxels in the human hippocampus using MVPA, suggesting that episodic-like memory traces are present and detectable within the human hippocampus. However, each episode in these studies differed along a variety of dimensions, including the identity of the people involved, the actions performed and the spatial contexts. It was therefore not possible to determine exactly how the event-like episodes were represented within the hippocampus, or precisely what aspect of the episodes was being detected by the MVPA classification technique.

The purpose of the current study was to apply similar MVPA methods to the study of highly overlapping episodic-like memories, in order to determine whether it was possible to detect unique, bound memory traces within the human hippocampus and elsewhere in the MTL. The overlapping information in the episodes was a critical aspect of this study, as it was important to ensure that no episode could be uniquely specified by any single element within it. In order to create such fully controlled stimuli, I filmed two brief action events against a green-screen background. Each event was then superimposed onto the same two spatial contexts, creating

four movie clips which included every combination of the two events and the two contexts (Figure 24). Each participant viewed the four movies prior to scanning, and then vividly recalled each one numerous times during high-resolution fMRI scanning. As the four episodes completely overlapped with one another in terms of their constituent elements, any successful differentiation of the four memories from patterns of activation must be due to the presence of unique, bound memory traces. If the hippocampus is exclusively involved in creating and maintaining distinct memory representations, it should be possible to decode highly overlapping episodic-like memories from the patterns of activity across voxels in the hippocampus, but not from other MTL regions.

5.2 Methods

5.2.1 Participants

Fifteen healthy right-handed participants (eight female, seven male) took part in the experiment (mean age 21.17 years, SD 2.18 years, range 18–25 years). All had normal or corrected-to-normal vision and gave informed written consent to participation in accordance with the local research ethics committee.

5.2.2 Stimuli

I filmed two brief action events against a green-screen background. Each event featured a woman carrying out a short series of actions (with a different woman in each event), each lasting 7 seconds. In the first, a woman

walked into shot, removed her jacket and placed it over her arm. In the second, a woman walked into shot, took out and put up an umbrella. These two events were then superimposed onto two different spatial contexts, creating four movie clips which included every combination of the two episodes and the two contexts (Figure 24). These stimuli ensured that the memories would be dynamic and episodic-like in nature, whilst being fully controlled in terms of the event content and spatial context of each memory.

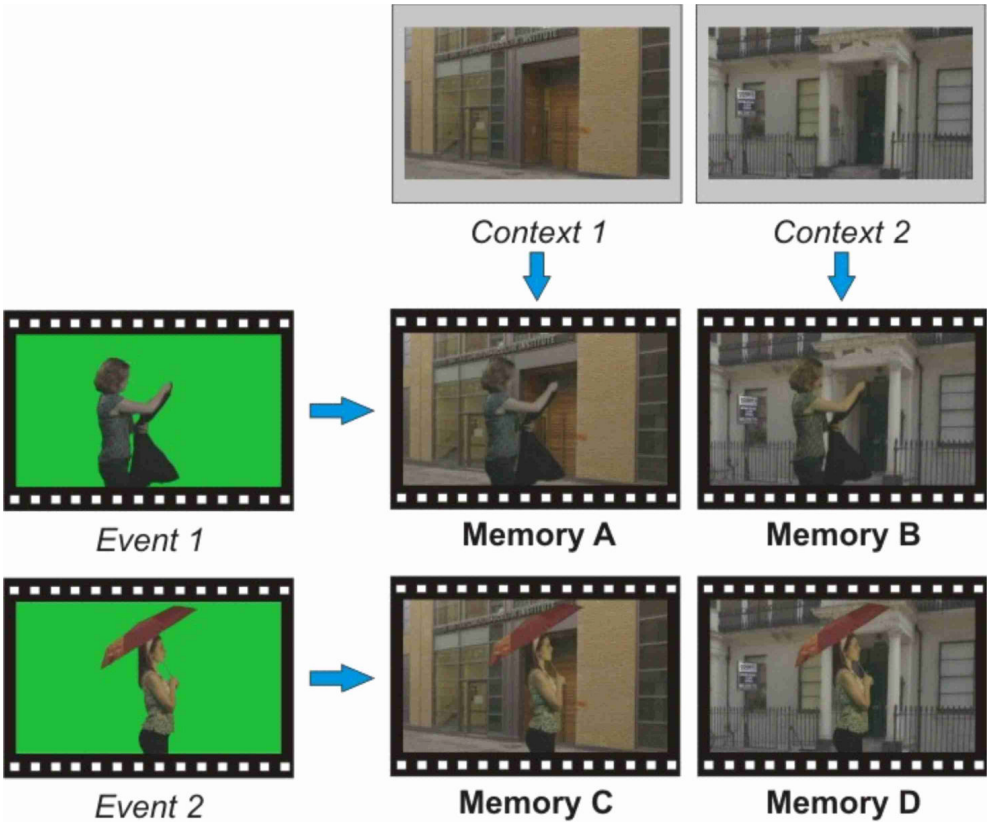

Figure 24. ***The movies***. *Two events were filmed against a green-screen background (left panels). The two events were superimposed on two different spatial contexts (see contexts in uppermost panels) in order to create four movies which included all four combinations of event content and spatial context (see panels Memories A-D). These stimuli ensured that the memories of them would be dynamic and episodic-like in nature, whilst being fully controlled in terms of the event content and spatial context.*

5.2.3 Pre-scan training

During a pre-scan training period, participants watched each of the four movie clips 12 times in total, and practised vividly recalling a movie after each viewing. This degree of training was necessary in order to ensure that participants were able to recall every memory consistently and accurately on every trial. To rule out any order effects in the initial presentation of the movie clips, participants viewed the clips initially in one of two different orders: in one case the first movie clip viewed contained the same spatial context as the second clip, but differed in terms of event content. In the second case, the first movie clip viewed contained the same event content, but differed in spatial context from the second clip. This order was counterbalanced across participants, with 8 participants in one cell and 7 in the other. For each result reported I tested for differences between these two counterbalanced cells, and no significant differences were found, demonstrating that order effects did not have a significant impact on the results.

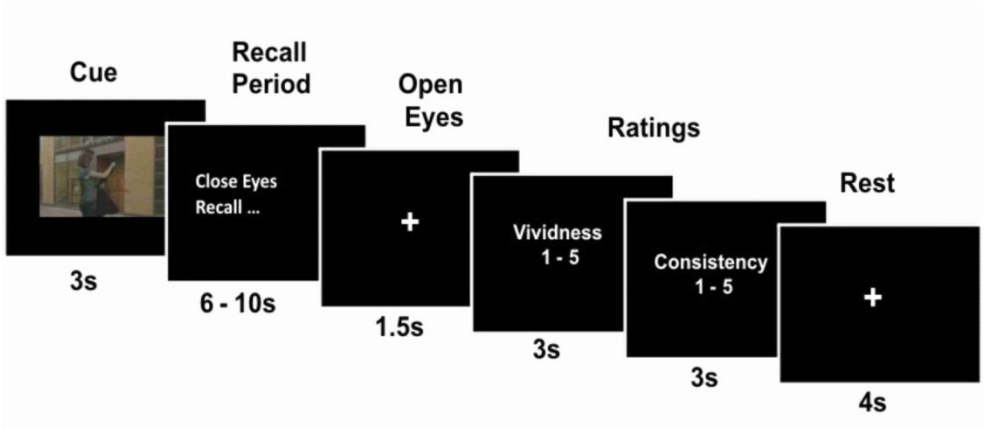

Figure 25. Timeline of a sample trial during fMRI scanning. *On each trial, one of the four episodes was cued with a still photograph taken from the movie. Following this cue, participants were instructed to close their eyes and recall the episode as vividly and accurately as possible, after which behavioural ratings of the recall experience were taken.*

5.2.4 Scanning task

Participants were scanned during recall of the four memories in a single scanning session. On each trial (Figure 25), a pictorial cue was presented for 3s indicating which of the four memories the participant had to recall. This cue was simply a still photograph taken from the relevant movie clip. Following this cue, participants were instructed to "Close Eyes" and "Recall", at which point they had to recall the relevant movie as vividly and accurately as possible. In order to ensure that the recalled memory approximated the original 7s length of the movie clip, the participant was required to press a button when they had finished recalling the clip (using a scanner-compatible keypad). If the button was pressed too soon (<6s) or they failed to push it within 10s, the participant would hear a tone, and a message would appear for 1.5s indicating that their recall had been too fast or too slow. Any such trials were excluded from the subsequent analysis. If the participant pressed the button within the allowed time, a fixation cross appeared on screen for 1.5s. Following this, the participant was required to provide ratings (3s allowed per rating) about the preceding recall trial using the five button keypad. First, they rated how vivid the preceding recall experience was (scale: 1 – 5, where 1 was not vivid at all, and 5 was extremely vivid). Second, they rated how consistent the recalled memory was with the other recall trials of that same memory (scale: 1 – 5, where 1 was not consistent at all, and 5 was extremely consistent). Any trials where a participant recorded a rating of less than 3 were excluded from the subsequent analysis. Following the ratings, participants rested for 4s before starting the next trial. In total there were 20 trials of each memory, presented in a pseudo-random order, whilst ensuring that the same memory was not

repeated twice or more in a row.

5.2.5 Post-scan debrief

After the scanning session participants completed a debrief questionnaire, assessing various factors relating to each memory. The specific questions are listed below:

How difficult was it to retrieve the memory from this cue? 1- 5, where 1 is very easy and 5 is very difficult.

How vivid, in general across the whole experiment, were the memories for this clip? 1- 5, where 1 is not vivid at all and 5 is extremely vivid.

Did you recognise the person featured in the clip?

Did you recognise the location featured in the clip?

How emotional did this clip make you feel? 1 – 5, where 1 is sad, 3 is neutral, and 5 is happy.

How much did this clip make you think about a real memory from your own life? 1 – 5, where 1 is not at all, and 5 is a lot.

How much did this clip make you think about yourself? 1 – 5, where 1 is not at all and 5 is a lot.

How much did you find yourself thinking about some sort of background story behind the clip? 1 – 5, where 1 is not at all, and 5 means you were thinking about a background story throughout.

How much did you find yourself trying to take the perspective of the person in these clips? 1 -5, where 1 is not at all and 5 is a lot.

How integrated did the memory feel, as a whole? 1 – 5, where 1 is not integrated at all, and 5 is extremely integrated.

I also asked the following questions regarding general experience throughout the scanning experiment:

Rate the degree to which you managed to keep your attention on the task all the way through the experiment, on a scale of 1-5, where 1 is poor attention and 5 is good attention all the way through.

How much were you aware of the commonalities between the different memories? 1 – 5, where 1 is not at all, and 5 is aware of them throughout.

Did you feel that you treated the four clips as distinct memories? Rate the overall distinctiveness from 1 – 5, with 5 being very distinct.

5.2.6 Image acquisition

All functional images were acquired using the high-resolution fMRI sequence that I described in Chapter 2. Field maps were acquired for distortion correction. T1-weighted MDEFT whole-brain structural scans were acquired for each participant after the main scanning session, with a whole brain T1-weighted 3D FLASH sequence (resolution 1 x 1 x 1 mm) acquired after the MDEFT sequence. In addition, high-resolution (0.52 x 0.52 x 0.5 $mm^{3)}$ T2-weighted structural images were acquired in a separate session on a 3T Trio scanner, as described in Chapter 2 (these are relevant for the analysis described in the next chapter). Four images were collected for each participant. These were then co-registered and averaged in order to improve SNR.

5.2.7 Image preprocessing

T1-weighted structural images were manually segmented into four regions (left and right): hippocampus, entorhinal cortex, perirhinal cortex, and posterior parahippocampal cortex using the ITK-SNAP software (Yushkevich et al., 2006, www.itksnap.org), according to the protocol described by Insausti et al. (1998). The anatomy of the hippocampus was further identified using Duvernoy (2005). All functional data were preprocessed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). The first six EPI volumes were discarded to allow for T1 equilibration effects

(Frackowiak et al., 2004). The remaining EPI images were co-registered to the T1-weighted structural scan, and then realigned and unwarped using the field maps (Andersson et al., 2001; Hutton et al., 2002). Each EPI volume was minimally smoothed with a 3mm FWHM Gaussian kernel. Each trial of interest was then modelled as a separate regressor, using a single boxcar, and was convolved with the haemodynamic response function. All excluded trials were collapsed into a single regressor of no interest for each memory separately. Movement parameters were included as regressors of no interest. Participant-specific parameter estimates pertaining to each regressor (betas) were calculated for each voxel. These parameter estimates were then transformed into t-values by dividing the beta estimate by the estimate of the standard deviation, as t-values have been found to produce more stable classification results (Misaki et al., 2010 - see Chapter 2). This preprocessing pipeline therefore produced a single t-value map for every accurately recalled memory trial during the functional session, and these data were used in all the classification analyses.

5.2.8 MVPA analyses

Decoding analysis of the defined regions of interest were all conducted using a linear support vector machine (SVM) classifier from the LIBSVM toolbox (Chang and Lin, 2011), in each case using a fixed regularization hyperparameter C = 1. Note that in this experiment, the added sensitivity from using t-values (see above, and Chapter 2) allowed me to use the whole ROI without feature selection. Therefore all analyses were applied to the whole ROI mask in each case, without using a feature selection step.

The first analysis was designed to determine whether it was possible to decode between the four overlapping memories from each of the four MTL regions (Figure 26A). For this analysis, a four-class classifier was applied to the data, using a leave-one-trial-out cross-validation procedure (Duda et al., 2001; Hsu and Lin, 2002). There was always one trial in the testing data partition, and the remainder were assigned to the training partition. Note that the exact number of trials differed across participants, as different numbers of trials were removed from each participant due to low ratings (see Scanning Task section above). Overall, this procedure produced an accuracy value for the region of interest based on the percentage of trials that were correctly classified. The set of accuracy values across the group of subjects was then tested against chance level of 25% (as there were four different memories) using a one-tailed t-test.

A second analysis was designed to test whether or not there was any common spatial information within any of the four MTL regions. In order to investigate this the classifier was trained to discriminate two memories which shared the same event content but differed in spatial context, using all recall trials for both memories. The classifier was then tested on the remaining pair of memories (Figure 26B). This analysis was performed in both possible directions (e.g. train on memories A vs. B, then test on C vs. D and then train on C vs. D and test on A vs. B), producing two accuracy values. These accuracy values were averaged to create a single representative accuracy for each region and participant. This set of accuracy values across the group of subjects was then tested against chance level of 50% (as there were two different memories in the test set) using a one-tailed

t-test. I also conducted a similar analysis to look for representations of common event content information, by training on memories A and C (where spatial context is exactly matched and the memories only differ in terms of event content) and testing on memories B and D, and vice versa.

Analyses were applied separately to the left and right hemisphere for each MTL region, and also to a set of combined masks which included all voxels from both left and right hemispheres for each MTL region. For every analysis and region, a comparison between the accuracy values in the left and right hemisphere was conducted using a paired t-test. None of these tests demonstrated any significant hemispheric difference, and therefore all results reported are based on the combined masks. The mean number of voxels contained within each of these four combined bilateral ROIs was: hippocampus (HC) = 1923 (SD 182.51), entorhinal cortex (EC) = 958 (146.68), perirhinal cortex (PRC) = 1317 (339.98), parahippocampal cortex (PHC) = 1167 (133.71).

5.2.9 Misclassification analysis

Misclassifications by the classifier could be one of three types: (a) incorrectly classified as a memory that shares the same spatial context (spatial misclassification); (b) incorrectly classified as a memory that shares the same event content (content misclassification); or (c) incorrectly classified as a memory which shares neither spatial context nor event content (orthogonal misclassification). In order to assess any potential biases on the four-class classification results, the proportion of misclassification trials falling into the three categories was calculated for

each participant. Comparisons of these misclassification rates were carried out using one-tailed t-tests.

5.2.10 Permutation testing

As this dataset was collected in a single functional session, I carried out a further set of analyses in order to ensure that the results reported were not due to non-independence of the training and test dataset. For each significant result reported above, a re-analysis was conducted using permutation testing (Nichols and Holmes, 2002; Etzel et al., 2009). Permutation testing also ensures that the results are not biased by an unbalanced classification design; as trials with low ratings were excluded from the analysis, the number of trials in each memory condition was not precisely matched for every participant. The permutation testing incorporates this unbalanced design, and provides an empirical null distribution of the data given that design. For each analysis, the classifier labels were randomly shuffled 1000 times, and for each shuffle an accuracy value was calculated in exactly the same way as described for the genuine labels. The real accuracy value was then compared to this permuted null distribution, and converted into a rank out of 1001 (consisting of the 1000 shuffled accuracies plus the real accuracy), where a rank of 1001 indicated that the real accuracy was greater than all the permuted accuracy values. The set of ranks for the group of participants was then compared to a chance level performance of 500.5 (middle point of 1001 ranks) using a one-tailed t-test. In each case, the permutation analysis confirmed the validity of the original results.

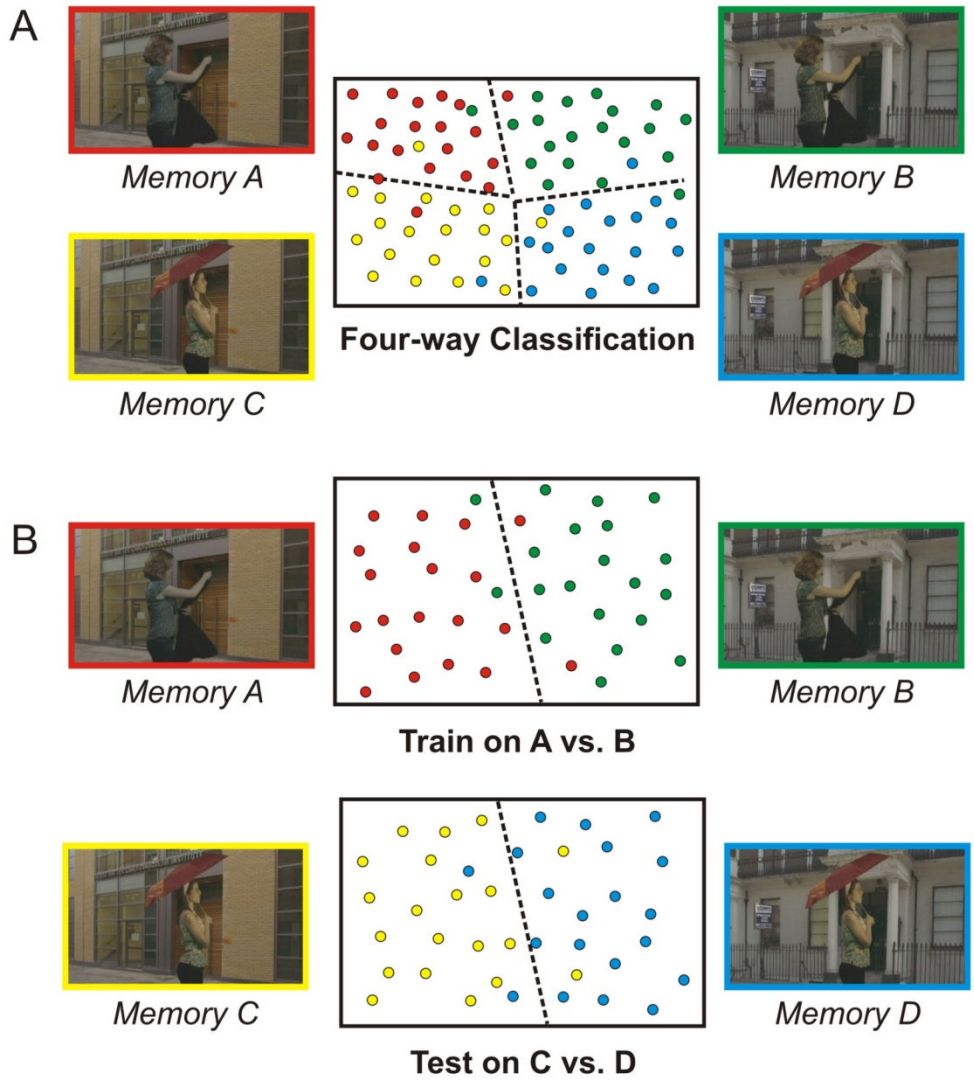

Figure 26. An overview of the decoding analyses*. (A) An illustration of the four-way classification procedure. The classifier was trained to find patterns of activated voxels for each of the four memories which best differentiated it from the other three memories. In this simplified schematic, each of the four memories is colour-coded, and each coloured dot represents the activity profile of a single recall trial projected into multi-dimensional space. The classifier was trained to find divisions within this space that best differentiated the activity patterns associated with each memory, here represented by the dotted lines. In this case each of these four regions is dominated by activity related to one of the four memories, demonstrating that the classifier has been able to find distinct patterns for each individual memory. (B) An illustration of the spatial context classification procedure. In this analysis I was interested in seeking information about spatial context that was common across different memories. In order to do this a classifier was trained to differentiate memories A and B, where the event content is exactly matched, and the memories only differ in terms of spatial context. If any spatial context information is present across pairs of memories, then the classifier that has been trained on A vs. B should successfully classify memories C vs. D, as the spatial contexts are exactly the same i.e. A and C share Context 1, and B and D share Context 2.*

5.2.11 Controlling for the number of voxels

As the total number of voxels differed between our regions of interest (see MVPA Analyses section), I ran a further control analysis to ensure that the difference in accuracy between the hippocampus and other MTL regions was not due to voxel number. In order to do this, I repeated the four-class analysis in each region 100 times, each time selecting a random set of 500 voxels. I then took the mean accuracy value across these 100 analyses for each participant, and compared this averaged accuracy against chance level accuracy with a one-tailed t-test for each ROI.

5.2.12 Examining the effects of smoothing

In line with previous studies of this kind, I applied 3mm of spatial smoothing to the functional images (see image preprocessing section above). The reason for applying this minimal level of smoothing is that even a small residual misalignment between the functional MRI volumes could have a negative impact on decoding analyses. This is particularly true for high-resolution data as in this study. The smoothing is assumed to stabilise the activity at each voxel, and mitigate against this kind of realignment issue. However, to investigate the effect of smoothing in this context, the hippocampal four-class analysis was repeated at three further levels of smoothing – unsmoothed, 6mm and 9mm.

5.3 Results

5.3.1 Behavioural results

Of the 80 memory trials during scanning, on average 10.2 (SD 10.48) were excluded due to low ratings. The mean vividness rating across all 80 trials was 3.8 (0.61) and consistency was 3.73 (0.66). The mean vividness rating for those trials included in the final analysis was 3.95 (0.61) and consistency 3.96 (0.59). Table 5 shows the numbers of trials included for each memory separately, along with the debrief ratings. The right-hand columns display the F and p values from a one-way repeated measures ANOVA, which demonstrate that there were no significant differences between the four memories for any of these ratings. This suggests that these extraneous factors did not drive the decoding results. Overall, participants were generally aware of the commonalities between the four episodes, with a mean rating of 3.13 (SD 1.3) out of 5. However, at the same time, they did manage to perceive the four memories as discrete episodes during retrieval, with a mean rating of 4 (SD 0.76) out of 5. Participants rated their ability to pay attention reasonably highly, with a mean rating of 3.5 (SD 0.63) out of 5.

Variable	Mean Scores (SD)				ANOVA	
	Memory A	Memory B	Memory C	Memory D	F	p
No. of trials included in each condition	15.40 (3.00)	16.53 (3.31)	15.67 (3.79)	15.33 (3.37)	0.4	0.75
How difficult was it to retrieve the memory?	2.43 (0.98)	2.07 (1.10)	2.20 (0.94)	2.60 (1.12)	0.79	0.51
How vivid was this memory?	3.47 (0.99)	4.07 (0.80)	3.93 (0.96)	3.80 (0.86)	1.21	0.31
How emotional did the memory make you feel?	3.00 (0)	3.00 (0)	3.00 (0)	3.00 (0)		
How similar was the memory to a real memory from your own life?	1.53 (1.06)	1.80 (1.26)	1.47 (0.64	1.60 (0.83)	0.33	0.81
How much did the memory make you think about yourself?	1.50 (0.94)	1.53 (1.13)	1.47 (0.74)	1.47 (0.74)	0.02	0.99
How much did you think about a background story?	1.73 (1.28)	2.30 (1.41)	2.40 (1.45)	2.50 (1.38)	0.93	0.43
How much did you take the perspective of the person?	2.30 (1.13)	2.20 (1.21)	2.60 (1.35)	2.20 (1.26)	0.35	0.79
How well integrated did the memory feel?	3.53 (0.64)	3.73 (0.88)	3.33 (1.18)	3.20 (1.15)	0.48	0.84

Table 5. Memory debriefing ratings. *A summary of the means and standard deviations of the debriefing scores (on a scale of 1 – 5, low –high) given for each of the four memories, as well as the mean number of trials included from each condition in the MVPA analyses after low rating trials were excluded. For the emotional response question, 1 = sad…3 = neutral…5 = happy. For each item, I tested for a difference between the four memories with a repeated measures one-way ANOVA, and the F and p values are presented in each case. None of the comparisons showed any significant difference between the four memories. Additionally, no participant recognised either the people or contexts in the movies.*

5.3.2 Four-class MVPA classification

A four-way linear SVM classifier was used to determine if it was possible to discriminate between the four overlapping memories from the activity across voxels in four regions of the MTL: the hippocampus (HC), the entorhinal (EC), perirhinal (PRC) and parahippocampal (PHC) cortices (Figure 27A). A significant level of decoding was found in the hippocampus ($t=1.90$; $p=0.04$) which was in contrast to the other regions, none of which supported significant levels of decoding (EC: $t=-1.09$, $p=0.85$; PRC: $t=1.55$, $p=0.07$; PHC: $t=0.31$, $p=0.38$; Figure 27B). Even when controlling for the number of voxels included in the analysis for each region (see Methods), only the hippocampus produced a significant level of decoding ($t = 1.8$, $p = 0.047$), while the other regions were not significantly above chance (EC: $t=-1.16$, $p=0.87$; PRC: $t=1.23$, $p=0.12$; PHC: $t=-0.11$, $p=0.54$). This result shows that in this extreme example of overlapping memories, where no single element allowed the differentiation of a memory, the hippocampus contained distinct representations of each individual memory.

When I examined the effect of different levels of smoothing, neither the 6mm nor the 9mm results produced accuracy values that were significantly above chance level performance. The unsmoothed data, on the other hand, resulted in above-chance accuracy (chance=25%, mean accuracy=45%). These results clearly demonstrate that in this particular study, the information is only detectable at a high spatial resolution, and that a level of smoothing above 3mm (that used here) significantly degrades the detectability of the multivariate information. This suggests that high-

resolution acquisition may be important for decoding individual memory traces, although this may depend on the type of memory traces being studied – here I investigated highly overlapping memories, which may require a more fine-grained level of multivariate information than memories that are more distinct from one another.

5.3.3 Spatial context MVPA classification

This study design allowed me to make further inferences about the specific informational content within the hippocampus. As every memory shared its spatial context with one other memory, I asked whether there was evidence of a common spatial context representation across such pairs of memories. In order to test this, a classifier was trained to differentiate memories A and B, where the event content is exactly matched, and the memories only differ in terms of spatial context (see Figure 26B). If any spatial context information is present across pairs of memories, then the classifier that has been trained on A vs. B should successfully classify memories C vs. D, as the spatial contexts are exactly the same i.e. A and C share Context 1, and B and D share Context 2. Only the classifier operating on the hippocampal voxels displayed successful decoding of the common spatial representation ($t=2.39$; $p=0.02$; see Figure 27C), with no significant decoding in the other MTL regions (EC: $t=-0.05$, $p=0.52$; PRC: $t=0.16$, $p=0.88$; PHC: $t=0.21$, $p=0.42$). These results demonstrate that, in addition to representing the four individual memories, the hippocampus also contained representations of spatial contexts held in common across different memories.

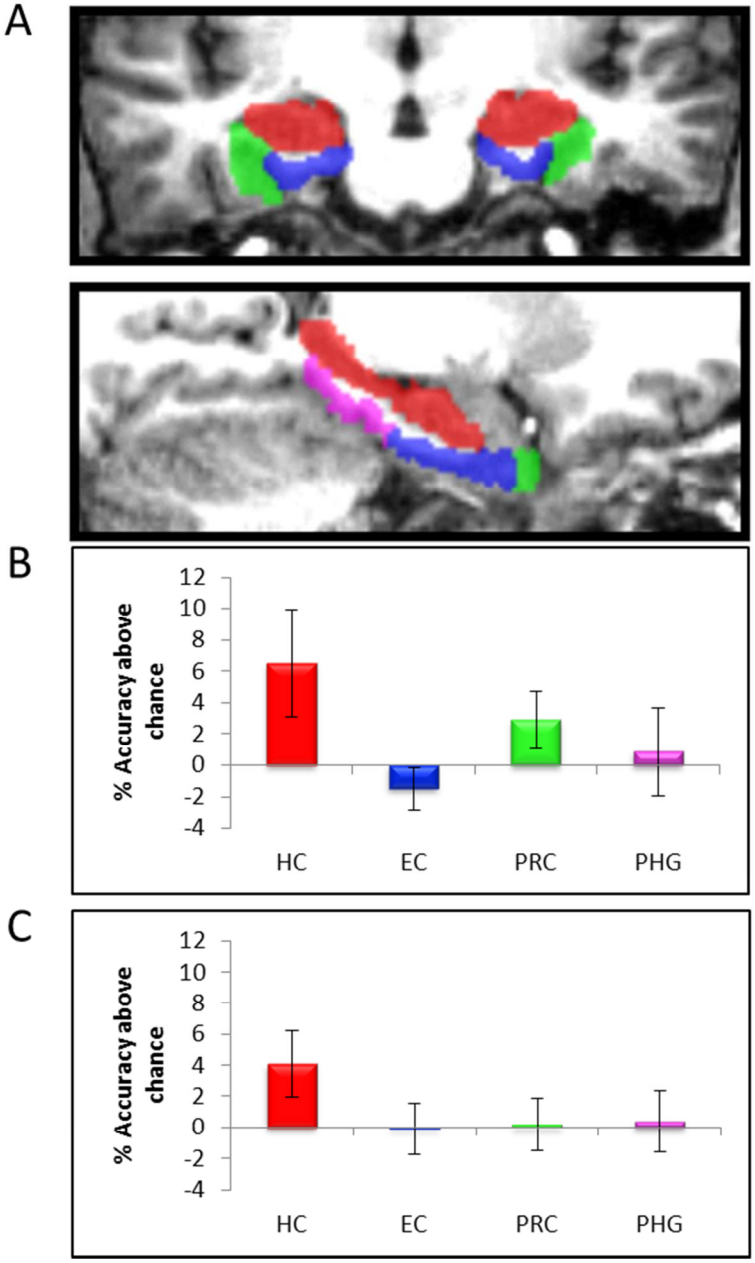

Figure 27. Summary of MVPA results*. (A) Segmented regions of interest in the medial temporal lobe of one of the participants shown in the coronal plane (upper panel) and sagittally (lower panel). The hippocampus (HC) is shown in red, entorhinal cortex (EC) in blue, perirhinal cortex (PRC) in green, and the parahippocampal cortex (PHC) in magenta. Group mean decoding results for each of the four MTL regions are displayed for (B) the four-way classification analysis, and (C) the spatial context classification analysis. Results are displayed as percentage above chance accuracy, with standard error bars. In both analyses, only the HC results are significantly above chance. Note that for both significant results, significance tests were repeated using a nonparametric permutation approach, and in each case the results remained significant.*

5.3.4 Event content MVPA classification

I also conducted a similar analysis to look for representations of common event content information, by training on memories A and C (where spatial context is exactly matched and the memories only differ in terms of event content – see Figure 26B) and testing on memories B and D, with no significant results in any of the four MTL regions (HC: t=0.51, p=0.31; EC: t=0.95, p=0.18; PRC: t=0.69, p=0.25; PHC: t=1.59, p=0.07). This suggests that while information relating to the fully bound memories and also the spatial contexts is relatively high and decodable in the hippocampus, information relating to event content alone is less so, at least in the four MTL regions that I examined.

5.3.5 Misclassification analysis

Given that the spatial contexts common to different memories were represented in the hippocampus, this raises an important issue regarding the initial analysis where all four memories were decoded using a four-way classifier. Theoretically, it would be possible to get above-chance decoding accuracy in the four-way analysis purely on the basis of spatial information, rather than specific information about each of the four memories, as four-way SVMs are based on a series of two-way classifications (Hsu and Lin, 2002). In order to rule out this explanation, I investigated the patterns of *misclassification* in the four-way analysis. On each trial, the classifier can either correctly classify the memory, or it can misclassify it. Misclassifications can be one of three types: (a) incorrectly classified as a memory that shares the same spatial context (spatial misclassification); (b)

incorrectly classified as a memory that shares the same event content (content misclassification); or (c) incorrectly classified as a memory which shares neither spatial context nor event content (orthogonal misclassification). If the four-way classification results are being driven by spatial information, then one would expect the misclassifications to be biased towards memories which share the same spatial context, and there should be a greater proportion of spatial misclassifications than either content or orthogonal misclassifications. Using a one-way paired t-test, I compared the number of spatial misclassifications against each of the other misclassification conditions. In neither case was the proportion of spatial misclassifications found to be greater (spatial>event misclassifications, t=-1.462, p=0.92; spatial>orthogonal misclassifications, t=-2.754, p=0.99). This demonstrates that the results of the four-way analysis were not driven by the representation of common spatial information, but instead genuinely reflect the representation of four distinct episodic memories within the hippocampus.

5.4 Discussion

Here I used MVPA decoding of high-resolution fMRI data to investigate the representations of highly overlapping episodes in the MTL during vivid recall. Of the four MTL regions tested, only the classifier operating on voxels in the hippocampus displayed a significant level of decoding between the memories. This shows that the hippocampus maintains distinct representations of episodic memories even when episodes are highly overlapping in terms of their constituent elements. Moreover, I found that,

in addition to representing the four individual memories, the hippocampus also contained representations of spatial contexts held in common across different memories.

While it is now well-established that the hippocampus is crucial for episodic memory (e.g. Scoville and Milner, 1957; Spiers et al., 2001; Burgess et al., 2002; Cipolotti and Bird, 2006), it is not known precisely how episodic memories are coded within the hippocampus. One influential account argues that episodic memories are stored as distinct memory traces within the hippocampus, even if those memories are highly overlapping in terms of their constituent elements (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). However, this hypothesis has not previously been empirically tested. In my previous investigations of episodic representations, I found that it was possible to decode episodic memories from fMRI BOLD activity in the MTL (Chapters 3 and 4), demonstrating that episodic representations are present and detectable within the hippocampus. However, the memories examined in those studies were all distinct, and differed along various dimensions such as spatial context, identity of actors, and the nature of the content. It was therefore not possible to determine exactly what information was being used by the classifier to decode the episodic memories, meaning that strong conclusions could not be drawn regarding representational coding of the episodic memories from those studies alone.

The current experiment permitted a conceptual advance over the previous studies, as the set of four episodes here were specifically designed so that no

individual memory could be uniquely identified by either spatial context or event content. This design ensured that successful decoding of all four memories could not be due to any single element such as spatial context or identity of the actor, as this information was shared across different memories. Instead, successful four-way decoding must rely on there being a distinct representation of each of the four memories above and beyond any general representation of spatial context and event content. Given that it was possible to successfully differentiate the four memories based on activity patterns across voxels in the hippocampus, this shows that the hippocampus does indeed contain a unique representational code for each memory, regardless of any shared components.

An important aspect of the study design was that it enabled me to make further inferences about the episodic representations within the hippocampus. In addition to its role in episodic memory, it has long been known that the hippocampus is critical for spatial representations and spatial navigation (e.g. O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978; Burgess et al., 2002; Hassabis et al., 2007, 2009). In a previous (standard, univariate) fMRI study the hippocampus was more active during recognition memory for both episodic and semantic information that included a spatial context compared with memories that did not explicitly require consideration of the spatial context (Hoscheidt et al., 2010). In that study, however, the contexts and content were not completely controlled, and it is not entirely certain whether the spatial context drove the hippocampal activations or an interaction between context and content.

By contrast, the memories in this experiment were completely controlled in terms of spatial context and event content. Given this, and given the multivariate approach to data analysis, I was able to ask a more challenging question - as every memory shared its spatial context with one other memory, was there evidence of a common spatial context representation across such pairs of memories in the MTL? The results of this analysis revealed significant levels of decoding within the hippocampus, but not the other MTL regions. This demonstrates that during episodic recall, in addition to representing the four individual memories, the hippocampus also contains a general representation of spatial context that is active during the recall of any memory sharing that spatial context. While this result is consistent with a wealth of evidence suggesting that the hippocampus is critical for spatial representations, as far as I am aware no previous study has isolated the representation of purely spatial information in this way during episodic recall. This provides a novel insight into spatial processing, and demonstrates that even during recall of internally generated, complex episodic-like memories, the hippocampus maintains a distinct representation of relevant spatial environments. It is interesting to note that I did not find any evidence for the presence of generalised spatial information within the parahippocampal cortex, a region documented to represent scene information in previous MVPA studies (Diana et al., 2008; Hassabis et al., 2009; Bonnici et al., 2011). One clear difference between this study and those previous studies is that here I examined spatial scenes that were generated as part of episodic-like memories. It is possible that spatial representations generated during episodic recall are more strongly represented within the hippocampus than in neighbouring regions. This will

require elucidation in future studies.

It is also interesting to note that while the current findings show that the hippocampus contains representations of the distinct bound memories and also the common spatial contexts, an analysis that explored whether the common event content was represented in any of the four MTL regions failed to produce significant decoding. This suggests that while information relating to the fully bound memories and also the spatial contexts is relatively high and decodable in the hippocampus, information relating to event content alone is less so, at least in the four MTL regions that I examined. This does not preclude such information existing elsewhere, beyond the partial volume used here. Moreover, it is possible that had I been able to look within hippocampal subfields, I might have detected evidence for content information, in line with previous studies such as Bakker et al. (2008). Notably, this failure to decode the event content from hippocampal activation bolsters the argument that the successful spatial context decoding analysis was indeed driven by the spatial properties of the context rather than any individual objects from the background (e.g. doors, railings). Each event also contained distinctly different objects (e.g. umbrella, jacket), and yet it was not possible to decode these events, suggesting that any signal regarding individual objects present within the hippocampus was not sufficient to drive successful classification performance.

One key question regarding these results is to what extent these representations reflect true episodic memories. Episodic memory is commonly defined as the memory for personally experienced events,

including details of the event along with the concomitant spatial and temporal context (Tulving, 1983, 2002). The events used in this study are not truly "episodic" under this definition, because each movie clip was presented 12 times during pre-scan training (to ensure that the memory representations were stable, which is necessary for MVPA), while genuine episodes are experienced only once. Nevertheless, I ensured that only those trials where there was vivid recall of the original movies were included in the analyses. It therefore seems likely that the core processes involved in the vivid recall of these episodic-like memories overlap considerably with those involved in episodic recall. Indeed similar levels of decoding within the hippocampus were achieved with genuine autobiographical memories (see Chapter 4).

In summary, these findings provide the first empirical demonstration that the hippocampus contains genuinely unique episodic memory traces, even in the presence of overlapping elements. Moreover, the results re-emphasize the fundamental role of the hippocampus in representing space (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978; Burgess et al., 2002; Hassabis et al., 2007, 2009), and demonstrate that spatial contexts are represented within the hippocampus during episodic recall. Together, this set of findings suggests that the hippocampus is capable of supporting at least two different types of representation - each memory has a unique representation, and at the same time spatial backdrops that are common to different memories are also represented in the hippocampus.

While the current results provide novel insights into the representational content of the hippocampus as a whole, important questions remain. For instance, one influential account of hippocampal function proposes that the hippocampus is able to store such overlapping episodes by orthogonalizing overlapping inputs into distinct memory traces, through the process of pattern separation (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). In the non-human hippocampus, region CA3 in particular has been implicated in the formation and maintenance of distinct, pattern separated representations (Leutgeb et al., 2004, 2007). At the same time, CA3 is also proposed to be critical for the process of pattern completion, whereby full episodic representations can be retrieved from partial cues (Lee et al., 2004; Vazdarjanova and Guzowski, 2004). It is therefore plausible that human CA3 could play a key role in the representation of unique episodic memories, but may also display pattern completion between the overlapping memories. Further MVPA studies investigating the representation of overlapping episodes within the subfields of the hippocampus may provide us with a means to bridge the gap between these theoretical neural computations and complex episodic memory.

6 Chapter 6

Decoding overlapping memories in the subfields of the human hippocampus

Precis

In the previous chapter, I showed that the hippocampus contains distinct episodic representations even when the episodes contain a high degree of overlap with one another. In the current chapter I used the data from the previous experiment and extended them, conducting a whole new set of analyses to probe the nature of the episodic information contained specifically within the subfields of the hippocampus. In order to do this, the participants' high-resolution, sub-millimetre structural MRI scans were manually segmented to identify the subfields. By comparing the representations contained within the different subfields, I was able to directly test some key predictions of the computational account of episodic memory, thereby bridging an important empirical gap in the literature.

6.1 Introduction

Theoretical models of episodic memory argue that the ability to form unique, distinct episodic representations depends on computations taking place within the subfields of the hippocampus (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). When we experience an episode, a process known as pattern separation, largely driven by the dentate gyrus (DG), leads to the formation of a unique, orthogonalized representation within region CA3. These distinct memory traces can be retrieved when a cue triggers completion of the original CA3 activity pattern (pattern completion), which in turn drives CA1, from where the entire distributed cortical memory trace can be reactivated (Marr, 1971;

Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Thus, the hippocampus as a whole, and region CA3 in particular, is implicated in the creation of distinct episodic memory traces, while allowing for flexible retrieval of those memories.

These computations theoretically allow the storage of many overlapping episodic representations which can each be retrieved independently and be experienced as distinct memories. Several recent studies from the rodent literature have shown that region CA3 displays a response profile consistent with a role in both pattern separation and pattern completion (Lee et al., 2004; Leutgeb et al., 2004, 2007; Vazdarjanova and Guzowski, 2004; Wills et al., 2005). More recently, functional MRI studies have produced evidence consistent with pattern separation processes in human CA3/DG in response to pictures of objects with graded levels of similarity (Bakker et al., 2008; Lacy et al., 2011). While this set of studies has provided empirical support for the existence of these computations, none has yet demonstrated a direct link between the theoretical models and complex episodic memory. The primary aim of this study was to bridge this empirical gap.

The theoretical models produce two clear, testable predictions about the episodic representations we would expect to see within the subfields. First, it is proposed that pattern separation processes lead to the formation of unique, distinct episodic representations within CA3, even when episodes share overlapping information. At retrieval, this unique memory trace is reactivated, which then directly activates the memory traces stored within CA1 (Rolls, 2010; O'Reilly et al., 2011). I therefore expect that both CA3

and CA1 will be preferentially involved in the representation of each unique memory trace, despite the high degree of overlap between episodes. Second, during retrieval of an episode, I expect that pattern completion processes occurring within CA3 will lead to the partial activation of any overlapping episodes. Notably the information present within CA3 should be dominated by the deliberately retrieved memory, but I nevertheless expect to see some evidence for activation of the overlapping memories as well.

As a secondary aim, I wanted to explore whether these low-level computations could be causally related to differences between individuals in the subjective perception of overlapping episodes. For instance, does the representational distinctiveness of episodic memories within any subfield have a direct relationship with the perceived distinctiveness of those memories during episodic recall? Indeed, at an even more basic level, is there any evidence that the physical structure of the subfields can have a causal relationship with subjective mnemonic perception?

In the previous chapter, I described a study that used MVPA to investigate the representation of overlapping episodes in the hippocampus as a whole. This study design presented an ideal opportunity to explore the representational properties of the hippocampal subfields in order to test extant theories of hippocampal function (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Firstly, it allowed me to investigate the amount of episodic information that each subfield represents about the unique episodic representations (with the prediction that CA3 and CA1 should contain more information than the

other subfields). Secondly, the nature of the design also allowed me to directly assess whether the recall of an episode leads to the activation of overlapping episodes, through pattern completion (region CA3 is predicted to display pattern completion effects). I therefore conducted a new set of analyses of the previous dataset, specifically focussing on the hippocampal subfields. In order to do this, sub-millimetre, high-resolution structural scans were acquired in a separate scanning session. These images were used to manually segment the hippocampus into its constituent subfields (Figures 28 and 29), and MVPA analyses were conducted within each subfield separately. Novel, model-based MVPA methods were used in order to allow the direct comparison of episodic information across the different subfields (Friston et al., 2008; Morcom and Friston, 2012).

6.2 Methods

6.2.1 Experimental design

The participants and experimental design are described in Chapter 5.

6.2.2 Image acquisition

Details are provided in the previous chapter and Chapter 2. As well as the fMRI scans, the key images for this analysis were the high-resolution (0.52 x 0.52 x 0.5 $mm^{3)}$ T2-weighted structural scans, as described in Chapter 2. Four scans were collected for each participant. These were then co-registered and averaged in order to improve SNR.

6.2.3 Data preprocessing

All neuroimaging and statistical analyses were conducted using SPM8. The first six functional volumes were discarded to allow for T1 equilibration (Frackowiak et al., 2004). The remaining functional volumes were spatially realigned to the first image of the series, and distortion corrections were applied based on the field maps using the unwarp routines in SPM (Andersson et al., 2001; Hutton et al., 2002). Each participant's whole brain MT FLASH structural scan was then co-registered to a mean image of their realigned, distortion-corrected functional scans. Following this, the high-resolution T2-weighted structural average was co-registered to the MT FLASH structural scan, bringing all images into alignment (this co-registration was performed prior to the manual segmentation of the subfields). Functional data were left unsmoothed for the decoding analyses so that information present across patterns of voxels across these much smaller regions (i.e. the subfields) could be detected. All data were analysed in the native space of each participant, using subject-specific ROIs.

6.2.4 Segmentation of the hippocampal subfields

In humans, in vivo examination of subregions within the hippocampus has proved difficult, but recent advances in high-resolution structural and functional MRI have begun to make it possible to localise fMRI BOLD activity to specific hippocampal subfields with greater confidence (e.g. Zeineh et al., 2000, 2003; Bakker et al., 2008; Carr et al., 2010; Suthana et al., 2011; Duncan et al., 2012). All fMRI studies reporting hippocampal subfield findings that have been published to date employed a standard

mass-univariate approach to data analysis. In this section I describe a novel protocol for the application of MVPA to the hippocampal subfields.

In order to do this effectively, I (1) wanted to include the whole hippocampus, given that functional differentiation within the hippocampus is now well-established (Moser and Moser, 1998; Maguire et al., 2000; Gilboa et al., 2004; Kahn et al., 2008; Fanselow and Dong, 2010; Poppenk and Moscovitch, 2011); (2) sought to separate, as far as possible, each individual subregion from the others, to examine their specific contributions. (3) While many studies report high inplane resolution in their MRI scans (e.g. 0.39x0.39mm – Zeineh et al., 2000), this is often acquired in thick slices (e.g. 3mm). The skewed resolution from non-isotropic voxels distorts delineation of subfields (making it particularly difficult in anterior hippocampal regions), and interpolating the data to the higher resolution provides only a best estimation. To circumvent these issues I acquired data with isotropic voxels. It should also be noted that in using MVPA, the use of unfolding and flat-mapping to visualise activation in the subfields (e.g. Zeineh at al. 2000) is not suitable because local patterns of activity among clusters of voxels get disrupted if data are projected from 3D to 2D flat maps (Carr et al., 2010). Examining the literature for methods of delineating subregions of the hippocampal formation, it is surprising how these apparently reasonable criteria are not easy to satisfy. Numerous methods are described, but the challenge of achieving a widely-accepted and completely satisfactory procedure is obvious.

While an exhaustive review of extant methods is beyond the scope of this chapter, I summarise here the main issues. First, many methods do not in fact examine subfields in the whole hippocampus. Some restrict their analysis to a few slices of the hippocampus (Mueller et al., 2007) or just 1cm of the structure (Mueller et al., 2010), others do not delineate subfields within the head of the hippocampus (Zeineh et al., 2000, 2003; Eldridge et al., 2005; Ekstrom et al., 2009; Suthana et al., 2009; Preston et al., 2010), or its tail (Zeineh et al., 2000, 2003; Eldridge et al., 2005), while others just focus on the body of the hippocampus (Yushkevich et al., 2010), or on one specific subfield (e.g. CA1, Bartsch et al., 2011) or ignore others (e.g. CA3, Moreno et al., 2007). Second, aside from consideration of whether the whole hippocampus is available for analysis, from the data that is acquired, only two studies report being able to delineate CA2 (Yushkevich et al., 2010; Malykhin et al., 2010). In both cases the scanners used had high fields (4T and 4.7T respectively), thus identifying CA2 with confidence likely remains beyond the capability of studies using standard 3T scanners. More seriously, most methods do not have sufficient resolution or contrast to separate CA3 from DG (Zeineh et al., 2000; Eldridge et al., 2005; Kirwan et al., 2007; Bakker et al., 2008; Ekstrom et al., 2009; Suthana et al., 2009; Carr et al., 2010; Mueller et al., 2010; Preston et al., 2010). Functional differentiation within the hippocampus, be that down its long axis (Moser and Moser, 1998; Maguire et al., 2000; Gilboa et al., 2004; Kahn et al., 2008; Fanselow and Dong, 2010; Poppenk and Moscovitch, 2011), or within the subfields (Lee et al., 2004; Leutgeb et al., 2004, 2007; Vazdarjanova and Guzowski, 2004; Wills et al., 2005; Rolls, 2010; O'Reilly et al., 2011) is well-established. Not being able to examine the anterior and posterior portions of the

hippocampus, or being unable to distinguish the roles of CA3 and DG, limits the scope of studies and the conclusions that can be drawn.

A third issue concerns how delineation is actually achieved. Most of the papers cited above manually segmented the subregions. This is very time-consuming and ideally involves at least two operators in order to test the reliability of segmentation. Two main automated procedures have been reported. Operating at 4T and with its main focus the evaluation of clinical scans, Yushkevich et al. (2010)'s 'nearly automatic' segmentation procedure was able to delineate CA1, CA2, CA3, DG and subiculum. Some manual delineation of the hippocampi before the automated procedure could operate was still required. While seeming to achieve accurate subfield segmentation, unfortunately, as noted above, it was not possible to identify subfields in the head and tail of the hippocampus, only in the body, currently limiting its utility of this approach outside of the clinical domain. The other automated procedure for segmentation of hippocampal subfields is available as part of the FreeSurfer analysis programme (Fischl et al., 2002, 2004). The initial development of this procedure, and the basis of its current implementation, is on the manual subfield segmentation of the right hippocampi of 9 individuals ranging in age from 22-89 years where data were acquired at high resolution (0.38x0.38x0.8mm) and averaged over five scans to achieve better SNR (Van Leemput et al., 2009). The definitions of the boundaries of the subfields are very different from other protocols (e.g. Carr et al., 2010; Malykhin et al., 2010; Yushkevich et al., 2010), and do not seem to correspond to delineations from previous studies or indeed from anatomical atlases such as the Duvernoy (2005). Instead the delineations were based on

geometrical rules. The authors provide no rationale for the use of these specific boundaries, and cite no previous references using a similar protocol. In addition, how accurately their procedures generalise to scans acquired with much less resolution and SNR (e.g. Hanseeuw et al., 2011; Teicher et al., 2012) is also untested.

It is evident that delineation of hippocampal formation subregions, a prerequisite for my research question, remains a substantial challenge (van Strien et al., 2012). I considered the automated procedures too incomplete (Yushkevich et al., 2010) or inaccurate (Van Leemput et al., 2009) for my purpose. Instead, with my colleagues (Bonnici, Chadwick, et al., paper in preparation) we devised the following protocol: using a standard 3T MRI scanner, data were acquired in the form of high-resolution T2-weighted structural scans (0.5mm isotropic voxels – see Chapter 2 for details) which allowed us to increase subfield boundary contrasts. This permitted manual subfield segmentation of the whole hippocampus including head and tail, and allowed the separation of CA3 and DG (CA2 could not be separated and was included with CA3) using the Duvernoy (2005) hippocampal atlas as a guide.

Manual segmentation of the hippocampal subfields was conducted using the ITK-SNAP software package (Yushkevich et al., 2006). The anatomical descriptions of the hippocampus and its subfields in Duvernoy (2005) was used to identify the subiculum, CA1, CA3 and DG in the high-resolution structural scans. The subiculum links the hippocampus to the entorhinal area and is located medially in the hippocampus. The division between the

subiculum and CA1 is marked with a change of contrast on the T2 images and is clear in the sagittal view. The CA1 subfield continues from the subiculum and ends once the curve (genu) of the Cornu Ammonis is reached. The curve itself is considered to be the CA2/CA3 region, which I will refer to as CA3. The division between CA1 and CA3 was identified with a change in contrast in the scan viewed coronally. When distinguishing DG from CA1 and CA3, the hippocampal sulcus provides a clear boundary, separating the DG from these two subfields (Figure 28).

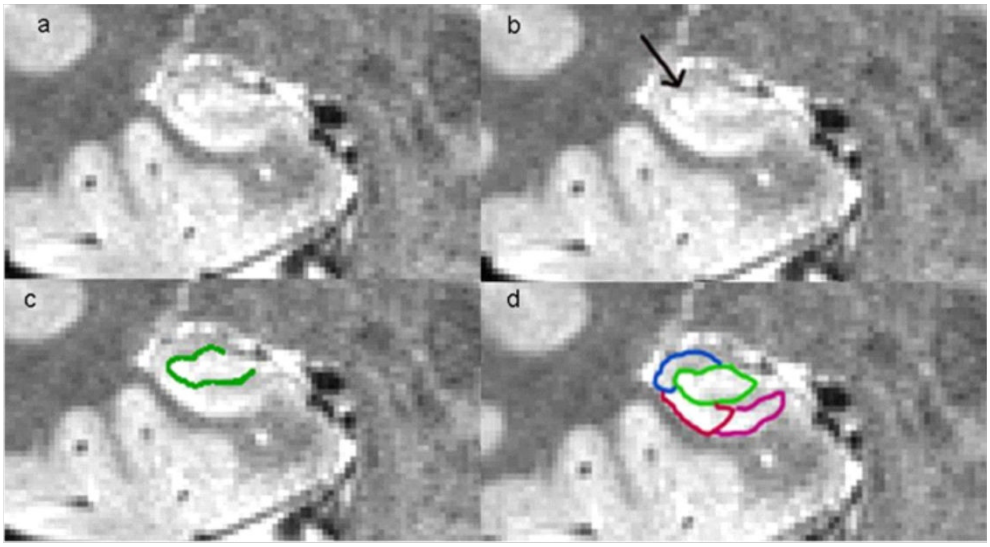

Figure 28. Subfield segmentation. *Note that these images are taken from a dataset used to design and validate this segmentation protocol, rather than the current data. (a) Original T2 image of hippocampus. (b) Arrow points to the hippocampal sulcus. (c) Green line indicates the hippocampal sulcus. (d) Outline of all subfields where magenta is subiculum, red is CA1, blue is CA3 and green is DG.*

Segmentation was performed in the coronal view, one subfield at a time starting with DG, then CA1, CA3 and finally subiculum. The starting point for segmentation was the slice where the body of the hippocampus emerged from the head of the hippocampus, distinguished as the point where the fimbria detaches from the head of the hippocampus, as described in

(Duvernoy, 2005), and working backwards through the body towards but not including the posterior tail of the hippocampus.

Once the segmentation of the body of the hippocampus had been completed, the view was rotated sagittally in order to segment the head and tail of the hippocampus. In this view segmentation started at the lateral edge, starting with only the CA1 subfield, and gradually progressing to defining CA1 (inferior) and CA3 (superior). Segmentation then proceeded medially through the hippocampus, segmenting CA1, CA3, and then DG as this subfield became visible. Subiculum was identified as a change in contrast between CA1 in the medial aspect of the hippocampus. At each point the axial view was used as reference point and to confirm that segmentation was correct. Once the sagittal segmentation was completed the view was rotated back to coronal to also confirm correct segmentation. These manual segmentations generated a set of masks for each participant for each hemisphere: CA1, CA3, DG and subiculum. The average amount of time taken to segment the subfields of one hippocampus was approximately 1 day.

A proportion of the hippocampi from the current study (four left, and three right) were segmented by a second trained operative in order to assess inter-rater reliability using the DICE (Dice, 1945) metric. The mean DICE scores for each subfield were as follows: CA3 – 0.68 (SD 0.03), CA1 – 0.78 (SD 0.03), DG – 0.75 (SD 0.02), subiculum – 0.58 (SD 0.03). These scores are similar to those reported by other methods (Yushkevich et al., 2010), indicating that the segmentations were reliable.

6.2.5 Behavioural variables

In order to investigate individual differences in the way that overlapping memories are perceived, I took some experiential ratings from each participant in a post-scan debrief session (see full details in Chapter 5). The ratings were designed to assess the subjective experience of recalling the memories during the scanning session. Two of these ratings are particularly important for my question of interest, as both tap into the subjective assessment of episodic distinctiveness. These two debrief items are included in Chapter 5, but I repeat them here:

Did you feel that you treated the four clips as distinct memories? Rate the overall distinctiveness from 1 – 5, with 5 being very distinct.

How much were you aware of the commonalities between the different memories? 1 – 5, where 1 is not at all, and 5 is aware of them throughout.

Thus, the first of these ratings assessed how subjectively distinct the participant perceived each memory to be during vivid recall in the scanner (which I subsequently refer to as "perceived distinctiveness"). The second assessed how aware they were of the commonalities between the different episodes during recall of each memory (which I subsequently refer to as "awareness of commonalities"). I investigated individual differences in perceived episodic distinctiveness using both of these behavioural variables in correlation analyses and a subsequent mediation analysis.

6.2.6 Decoding Analyses

6.2.6.1 Multivariate Bayes

For this experiment I used a Bayesian model-based decoding method called Multivariate Bayes (MVB) for all decoding analyses (Friston et al., 2008; Morcom and Friston, 2012). An MVB model maps multivariate voxel responses to a psychological target variable (e.g. individual memories), using a hierarchical approach known as Parametric Empirical Bayes. MVB uses the same design matrix of experimental variables used in a conventional SPM analysis (see Chapter 2). When a decoding contrast is specified, a Target variable X is derived from this contrast, after removing confounds. The multivariate voxel activity provides the predictor variable Y, which the MVB model will try to fit to X, ultimately producing a log model evidence, or Bayes factor for that model.

It is possible to specify priors on the pattern of voxel weights in an MVB design, and in this case I used a sparse prior, as the distribution of episodic representations is expected to be sparse (Rolls, 2010; O'Reilly et al., 2011). The model evidence for the sparse model is always compared to a null model, which assumes no mapping between the data and psychological variable, and the resulting log Bayes factor is the result of a model comparison between the sparse and null models. This can be considered as a measure of the mutual information between the multivariate data and the psychological variable. By explicitly modelling the mutual information in this way, MVB is potentially more sensitive than other decoding approaches such as support vector machines. Furthermore, because the multivariate data from a region is formulated as part of the model in an MVB design, it

becomes possible to directly compare information across different regions, as this now reduces to a model comparison (Friston et al., 2008). As noted in Chapter 3, comparing decoding results across different regions is problematic with most MVPA methods (Diedrichsen et al., 2011). Given that the key question in this experiment involved the explicit comparison of decoding results across the hippocampal subfields, this property of MVB was advantageous, and the main reason for employing it here.

For the MVB analyses I first set up an appropriate SPM design matrix. I created a single regressor for each individual memory, where every recall trial for that memory was modelled with a boxcar function covering the entire length of the recall period. Movement parameters were included as regressors of no interest. For all analyses, the log Bayes factors were treated as summary statistics, and used in classical statistical tests (Morcom and Friston, 2012). For all MVB models and each subfield, I tested for significant differences between the hemispheres, and found none. I therefore averaged the information measures across the hemispheres, and all analyses reported here are based on these pooled measures.

6.2.6.2 MVB models

In order to investigate the presence of unique information about each individual episode, I fitted four MVB models, one to each individual memory regressor (i.e. one model for Memory A, one regressor for Memory B etc. – see Figure 29). The log Bayes factors were averaged across these four models, creating a single summary measure of information about unique episodic information.

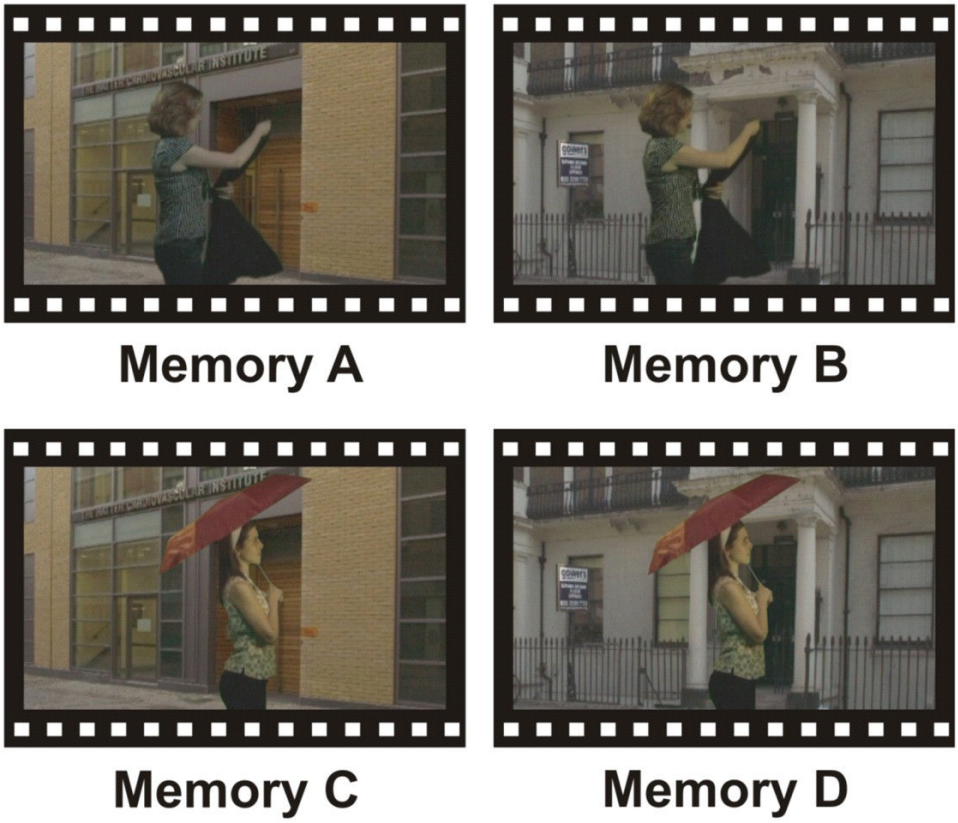

Figure 29. The movies*. Two events were filmed against a green-screen background. The two events were superimposed on two different spatial contexts in order to create four movies which included all four combinations of event content and spatial context (see panels Memories A-D). These stimuli ensured that the memories of them would be dynamic and episodic-like in nature, whilst being fully controlled in terms of the event content and spatial context.*

To investigate the presence of shared information across overlapping memories, I explicitly modelled two memories at a time. Each memory had another memory that shared spatial background, another memory that shared event content, and one that had no shared elements. For example, Memory A shared the spatial background with Memory C, shared the event content with Memory B, and shared no elements with Memory D (see Figure 29). I ran a separate MVB model for every pair of memories, and then averaged across the log Bayes factor for each type of pair (spatial

overlap, event overlap, no overlap). No significant differences were found between the spatial and event overlap models, so the log Bayes factor was averaged across these conditions in order to create a single measure of overlapping information.

6.2.6.3 Adjusting for subfield size

It is possible to increase the amount of measurable information in a region simply by providing more informative voxels. It is therefore important to ensure that any informational differences between regions cannot simply be attributed to differences in size. For each participant I calculated the relative size of each subfield (subfield size/total hippocampal volume), and used this to calculate the expected information in each region if the differences were completely explained by size. To do this I multiplied the total information summed across the subfields by the relative size of each subfield. Next I subtracted this size-predicted information from the measured information in each region to create a differential. Finally, this differential was added to the mean information across the four subfields. This therefore created an adjusted information score which reflected the amount of information that is not predicted from size alone, while preserving the total amount of information across the subfields. I used the adjusted scores for each MVB model, and these adjusted scores are reported and used for further analysis throughout the chapter.

6.2.6.4 Pattern completion measure

It is important to note that the log Bayes factor for the overlapping MVB model reflects two possible sources of information. First, there could be information that is shared across the memories, consistent with a pattern completion process. However, there could also be information about the two memories independently. Given the multivariate nature of MVB, it is perfectly possible that two independent sources of information can be detected, thus leading to a high model evidence (log Bayes factor). In order to remove this confound, I used the non-overlapping model as a baseline (i.e. the model where two non-overlapping memories are modelled together, e.g. memory of movies A and D – see Figure 29). As this model has no shared components, it can only detect information about two independent memories. I subtracted this baseline information from the overlapping information to create a pure measure of pattern completion information, and this is the measure reported in the results section.

6.2.6.5 Model-fit generalization

I conducted two further analyses based on the output of the MVB models. The predictor variable Y for each model represents the optimal predicted response of that memory based on the underlying pattern of multivariate activity. It is possible to fit this predictor variable Y against the original target variable X in order to assess the fit of that model using a linear regression. The parameter estimate (beta) from this analysis will give an indication of how well the multivariate response predicts the memory. Interestingly, it is also possible to fit predictor variable Y of a memory (e.g. memory A) to target variable X of other memories (e.g. memory B). The

resulting parameter estimate from this analysis will inform about how well the modelled multivariate response of memory A also predicts memory B. I used this general principle to conduct the analyses described in the next two sections.

6.2.6.6 Model specificity

The first set of MVB models modelled the responses to each individual memory. However, it is possible that a model may be picking up on information that is shared across different memories, especially in the case of overlapping memories. In order to assess the specificity of each of these models, I used a linear regression to fit predictor variable Y against the original target variable X for each memory, thereby calculating a set of "modelled memory" betas. I then fitted predictor Y of each memory against target variable X of each other memory, to assess how much the model generalized to the other three memories. In order to generate a score of episodic specificity (i.e. an index of how specific the episodic representation was to an individual memory), I then subtracted the maximum of the three generalized betas (thereby using a conservative approach) from the "modelled memory" betas. If there was a significant amount of episodic information that was specific to the modelled memory, this measure should be significantly greater than zero. Note that this episodic specificity score was used in the subsequent individual differences analyses.

6.2.6.7 Model Generalization

I used the same model-fitting approach to provide a second measure of pattern completion. Here I took the betas from generalizing across each pair of overlapping memories (as described above), and those from each pair of non-overlapping memories. I then directly compared the two sets of betas using a paired t-test. If there is any evidence for pattern completion, we would expect to see evidence for significantly greater generalization across the overlapping pairs.

6.2.7 Mediation Analysis

To formally test for the presence of a mediation relationship in the individual differences data, I applied a mediation analysis. Mediation analysis is a form of path analysis, where the (linear) causal relationship between three variables is assessed. Specifically, a mediation analysis is interested in determining whether there is any evidence that X has a causal effect on Y *via* a mediating variable M. In other words, rather than X directly causing Y, the causal relationship is encapsulated by the fact that X causes M, which in turn causes Y. This causal path is known as the *indirect path*, often labelled as *ab*. If the indirect path has a statistically significant effect on Y, then one can conclude that there is a significant mediation effect present within the data. A bootstrap method (Preacher and Hayes, 2004, 2008) was used for this mediation analysis using 10,000 permutations, and it was implemented within MatLab using the BRAVO toolbox (https://sites.google.com/site/bravotoolbox/).

6.3 Results

6.3.1 Distinct episodic information

The first decoding analysis investigated the amount of unique information about each memory contained within each subfield during vivid recall (Figure 30b). The first result to note is that all four subfields contained significant levels of information (assessed using t-tests against zero; for all four subfields, $t > 16$, $p < 0.000001$), demonstrating that unique information is present throughout the hippocampus. However, as hypothesised, I found that regions CA3 and CA1 contained significantly more distinct episodic information than the DG and subiculum (assessed using an F-test: $F = 10.37$, $p = 0.0062$). Thus, the information was not evenly distributed across the hippocampus, but instead was biased towards regions proposed to be particularly important for the storage and retrieval of unique episodic representations (Rolls, 2010; O'Reilly et al., 2011).

I conducted a further analysis to ensure that the modelled information in the subfields was indeed specific to each individual memory, rather than generalizing across the overlapping memories. In order to do this, I fitted the decoding model of each specific episode to the data from each of the four episodes. The resulting betas were used to derive a measure of episodic specificity, which indexed the degree to which the episodic information was specific to the individual memories (see Methods). If the information was specific to the individual episodic memories, then this measure should be significantly greater than zero. To test this, I applied a t-test against zero to the episodic specificity scores in each of the four subfields, and found that

all four regions contained information that was specific to each unique memory trace, despite the high degree of overlap across the episodes (for all four subfields, $t > 7.5$, $p < 0.00001$). Consistent with the previous analysis, however, regions CA3 and CA1 were found to contain a significantly greater degree of episodic specificity than the DG and subiculum (assessed using an F-test: $F = 35.7$, $p = 0.00003$).

6.3.2 Episodic pattern completion

The second decoding analysis investigated information that was shared across overlapping episodes (see Methods). If pattern completion is causing the partial activation of overlapping episodes, then we would expect to find evidence for this shared information specifically within CA3. Using a one-way t-test, this is precisely what I found (Figure 30c), with significant generalized information within CA3 ($t = 1.79$, $p = 0.048$), but not in any other subfield (CA1: $t = -1.59$, $p = 0.95$; DG: $t = -0.71$, $p = 0.75$; Sub: $t = -0.36$, $p = 0.64$). Furthermore, a repeated-measure ANOVA revealed a significant effect of subfield ($F = 3.16$, $p = 0.034$), and post-hoc paired t-tests (one-way) clearly show that this effect was driven by CA3 containing significantly more information than both CA1 ($t = 2.31$, $p = 0.018$) and the subiculum ($t = 1.82$, $p = 0.045$), with a trend towards significance in the comparison with the DG ($t = 1.69$, $p = 0.057$). Thus, this result clearly supports the hypothesis that CA3 is particularly involved in pattern completion.

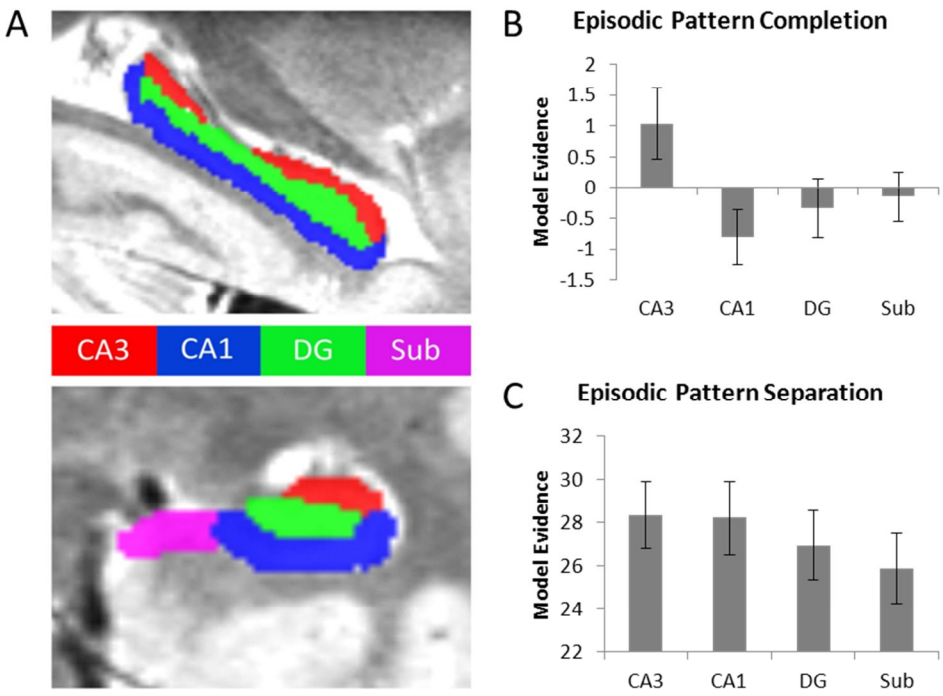

Figure 30. Episodic information in the hippocampal subfields. *(A) Manually segmented subfields of the right hippocampus in one example participant. The regions are displayed in the sagittal view in the top picture, and the coronal view in the bottom picture. (B) The top graph displays the results of the individual memory decoding analysis, with the model evidence (log Bayes factor) displayed on the y axis. These distinct episodic representations are significantly stronger in the CA fields than the other two subfields. (C) The bottom graph displays the results of the pattern completion analysis. The y axis displays the log Bayes factor. CA3 is the only region displaying evidence of episodic pattern completion.*

I used a second, independent method to provide further evidence for episodic pattern completion within region CA3. This analysis involved using the model for each individual memory to predict the data for each overlapping memory (see Methods), and this analysis demonstrated that overlapping memories could be predicted significantly better than chance ($t = 3.42$, $p = 0.0041$). Note that this is a strong test of pattern completion, as in this case each model has never "seen" the data from the overlapping memories, which therefore provides an independent test dataset akin to

classical MVPA classification analyses (see Chapter 2). Together, these two sets of results provide clear evidence of pattern completion within region CA3 during vivid recall of episodic memories.

6.3.3 Individual differences in episodic representation

These two sets of results demonstrate that region CA3 contains representations of both the unique episodic representation and the overlapping episodes during retrieval. This raises the interesting question of whether there are individual differences in the extent to which each of us is able to keep overlapping representations distinct within CA3. Is it possible that differences in the underlying representations could lead to qualitative differences in the subjective perception of overlapping memories? Is it further possible that the anatomical size of CA3 itself could have a direct causal relationship with our subjective mnemonic perceptions?

In order to address these questions, I investigated the relationship between the two ratings of subjective episodic distinctiveness ("perceived distinctiveness" and "awareness of commonalities") and (a) the distinctiveness of the episodic representations contained within CA3 (using the episodic specificity measure), and (b) the relative size of CA3 (CA3 size/total hippocampal size) for each participant. To do this I used Spearman's rank correlations (which is more robust to outliers than the Pearson correlation coefficient) between each pair of variables. This revealed a strong negative correlation between episodic specificity and the subjective awareness of commonalities (Spearman's rho = -0.76, p = 0.0011).

Thus, the more distinct the CA3 information about each episodic memory, the less the participants were subjectively aware of the overlap between the memories. This therefore suggests a direct causal mapping between CA3 informational content and subjective mnemonic perception (Figure 31). None of the other regions displayed any significant relationship with either subjective variable, even using a liberal threshold of $p = 0.05$, demonstrating that this effect is specific to CA3.

Interestingly, I found that the physical size of CA3 also showed a strong negative correlation with the awareness of commonalities (Spearman's rho = -0.71, $p = 0.0032$). Again, this correlation was specific to CA3, and no other regions showed any correlation with the subjective ratings. This tells us that not only is there a specific relationship between CA3 information and mnemonic perception, but we can also predict, with a high level of accuracy, the subjective distinctiveness of overlapping memories solely from the anatomical size of CA3.

These two correlations were specific to the “awareness of commonalities” subjective rating, and were not present with the “perceived distinctiveness” rating. It is possible, therefore, that the two ratings indexed two subtly different types of subjective experience, and that the former rating specifically related to CA3 anatomy and function. However, the range and standard deviation of ratings were much lower for the “perceived distinctiveness” (range = 2, SD = 0.76) measure than the “awareness of commonalities” measure (range = 4, SD = 1.30), despite both ratings being given on a scale of 1 – 5. This makes it more likely that the increased

variance in the “awareness of commonalities” measure provided more scope for detecting meaningful correlations than the “perceived distinctiveness” measure.

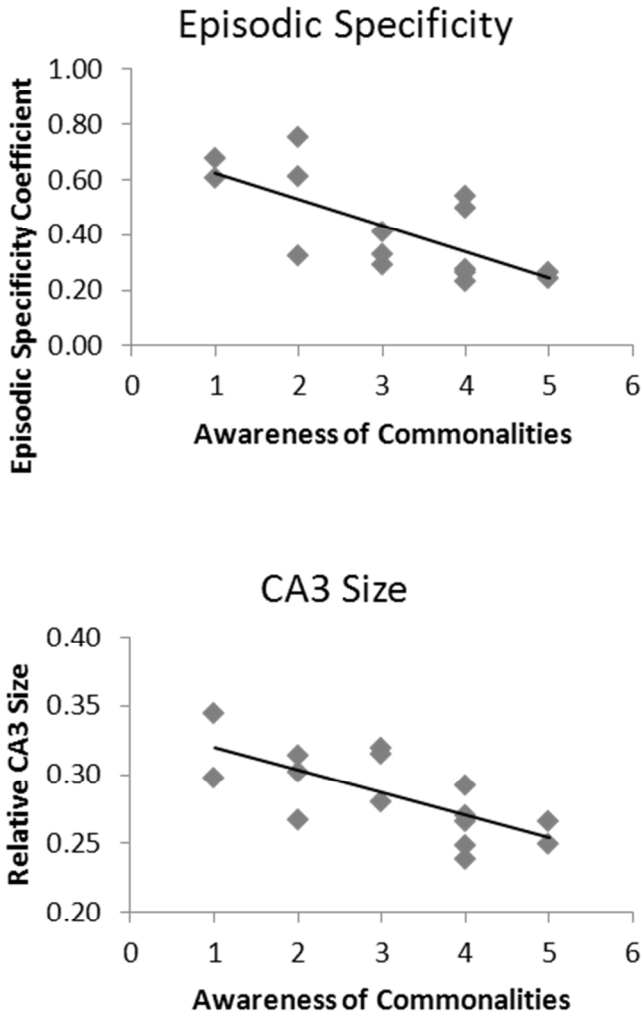

Figure 31. Correlations with individual differences in awareness of commonalities. *The top figure plots the specificity of episodic information within CA3 (y axis) against the subjective awareness of the commonalities across memories during episodic recall (x axis). The bottom figure plots the size of CA3 relative to the whole hippocampal volume (y axis) against the subjective awareness of the commonalities across memories during episodic recall (x axis). Together these correlations demonstrate a striking correspondence between CA3 information and structure, and individual differences in the subjective perception of episodic memory. The r values are from a Spearman's rank correlation in each case.*

6.3.4 Mediation analysis

The fact that both the anatomical size of CA3 and the episodic specificity within CA3 show the same negative correlation with subjective awareness of commonalities suggests a model of causality leading from CA3 anatomy to mnemonic perception. Specifically, I propose that increases in the physical size of CA3 lead to increased episodic distinctiveness within CA3, which then has a direct influence on the subjective distinctiveness of the recalled memories. To formally test this model, I applied a mediation analysis using a bootstrap method (Preacher and Hayes, 2004, 2008) with 10,000 permutations. This analysis revealed a significant effect via the indirect, mediation path ($p = 0.026$), thus providing empirical support for this model of causality. For the full set of path coefficients from this mediation analysis, see Figure 32.

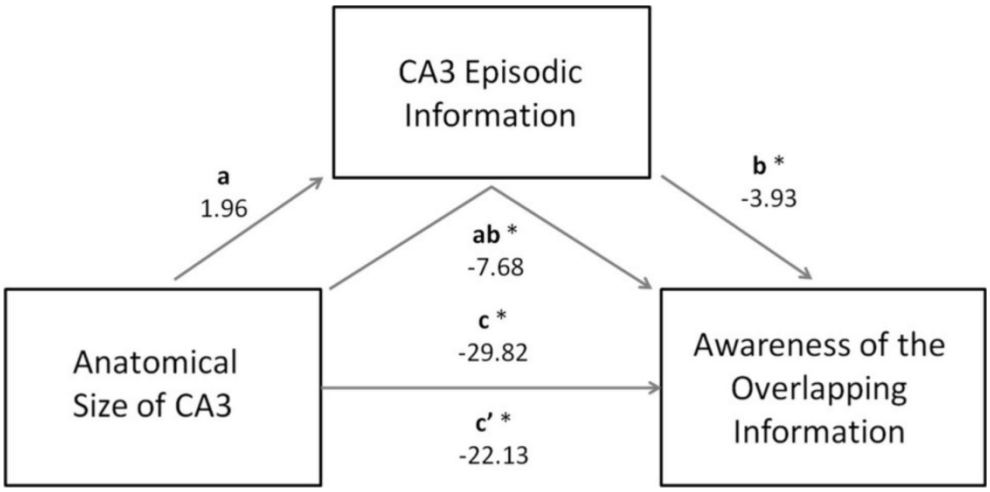

Figure 32. Mediation Analysis. *I tested for a causal relationship between differences in the anatomical size of CA3 and subjective episodic distinctiveness, mediated by the quality of episodic information contained within CA3. Path* ***a*** *indicates the effect of CA3 size on CA3 information. Path* ***b*** *indicates the effect of CA3 information on the awareness of*

*commonalities across the memories, after controlling for the effect of CA3 size. Path **c** is the total effect of CA3 size on awareness of commonalities. Path **ab** is the path of interest, and is known as the indirect path. It is simply the product of paths **a** and **b**, and directly assesses the mediation effect. Path **c'** is the direct path, and is calculated by subtracting **ab** from **c**. Stars indicate those paths which are considered to be significant using a bootstrap method. The critical result is the significance of the indirect path, which demonstrates the presence of a mediation effect.*

6.4 Discussion

Here I have provided the first experimental evidence in support of two key predictions from computational models of the hippocampus. First, during the retrieval of highly overlapping episodes, both CA3 and CA1 show evidence for the representation of distinct episodic memory traces. Second, I demonstrate that overlapping memory traces are concurrently activated within CA3 through a process of pattern completion, even though the information as a whole is dominated by the explicitly retrieved memory. Together, this set of results provides a vital bridge between computational theory and episodic memory, and allows us to make a stronger claim that such processes may indeed form the core neural mechanisms underlying human episodic memory. By so doing, it may now be possible to bring the study of episodic memory onto a more quantitative and rigorous theoretical footing.

To date, the dominant neurobiological accounts of episodic memory have been descriptive, cognitive process accounts. Two of the major theories state that the hippocampus is crucial for episodic memory (e.g. Squire, 1992; Nadel and Moscovitch, 1997; Squire et al., 2004; Moscovitch et al., 2005; Winocur and Moscovitch, 2011), without providing a fully detailed

description of the proposed neural mechanisms underlying this role (I should note that this was never the intention of either theory, but it is nevertheless a limitation). Alternative theories instead argue that the hippocampus is crucial for episodic memory due to its role in spatial (Andersen et al., 2006; O'Keefe and Nadel, 1978), relational (Cohen and Eichenbaum, 1993; Eichenbaum, 2000, 2004), or scene construction (Hassabis and Maguire, 2007, 2009) processes. Again, these high-level cognitive theories do not provide detailed accounts of the neuronal representations and computations involved. While each of these theories has contributed much to the field, they ultimately do not provide a fully mechanistic account of episodic memory. By demonstrating a link between episodic memory and existing computational theories, I show that these theories offer a viable additional framework for understanding complex episodic memory in the human brain. Thus, these theories in combination with further in-depth study of hippocampal subfield processing may allow us to start mapping out the detailed neural underpinnings of episodic memory in terms of neural representations, and the computations that are performed on those representations.

Before I move on to discuss the individual differences results, it is worth discussing one important point regarding the above set of findings. Here I show that both CA1 and CA3 are particularly involved in the representation of distinct episodes during episodic recall. This may at first glance appear to be somewhat at odds with previous studies showing that the DG/CA3 is crucially involved in the representation of distinct object representations, and not CA1 (Bakker et al., 2008; Lacy et al., 2011). However, there is an

important distinction between these two studies (in addition to the fact that here I investigate complex episodic representation as opposed to single objects). In this study I was explicitly interested in investigating the representation of well-learned episodes during retrieval, whereas the previous two experiments investigated novelty and adaptation responses to novel and lure stimuli. Thus, the previous experiments were effectively investigating encoding-related activity. This is a critical distinction, as the computational models explicitly predict that connections between CA3 and CA1 adapt rapidly, such that representations within CA1 should be learned very quickly. The CA1 representations then act to (a) stabilise the CA3 memory trace (Treves and Rolls, 1994; Rolls, 2010) and (b) provide a mapping between CA3 representations and the output structures of the hippocampus, thereby allowing reliable retrieval of distinct memory traces without catastrophic interference (McClelland et al., 1995; O'Reilly et al., 2011). Thus, as the study presented here investigated previously-learned representations, we would expect the CA1 representations to already be stable, and we would predict precisely the pattern of results that was found. If, however, we were to study the representations of similar overlapping episodes during initial encoding, it is possible that we would observe a pattern of results that is more similar to the previous studies, with distinct representations within the DG and CA3, and not within CA1 (Bakker et al., 2008; Lacy et al., 2011).

Another result that requires mention concerns the lack of evidence in this analysis for a specific representation of spatial context, which I reported in the previous chapter. Instead, I found evidence for pattern completion

effects within CA3 for both the spatial context and the event content, with no difference between the two types of information. What factors can explain this discrepancy? It is possible that when collating information across the whole hippocampus, as in the previous study, there may be a slight bias towards spatial context information that is not apparent within the individual subfields. In other words, while CA3 shows no significant difference in itself, it might be that there are subtle spatial context signals present across all the subfields that, when combined, provide enough information for the whole-hippocampus classifier to detect. Whether or not this is the case is not clear from the current dataset, and future studies will be required to clarify this issue.

In addition to the group-level results described above, I also investigated neural correlates of individual differences in the perceived distinctiveness of the overlapping episodes. This analysis revealed that the distinctiveness of the episodic representations measured within subfield CA3 showed a strong negative correlation with the subjective awareness of commonalities across the overlapping memories. Thus, the distinctiveness of overlapping memory traces within CA3 map directly onto individual differences in how subjectively distinct the memories appear to be. Notably, this effect was specific to region CA3, and was not apparent in any other hippocampal subregion. This specificity is entirely consistent with the proposed role of region CA3 in both pattern separation and pattern completion.

Indeed, one previous study has found evidence of a link between CA3/DG functional and structural measures and individual differences in object

discrimination in a group of older adults (Yassa et al., 2011). However, the work I present here is the first to demonstrate that informational content within region CA3 correlates directly with subjective experience of episodic recall. This result echoes the proposal that, during retrieval, there is a constant tension within CA3 between the activation of distinct representations, and pattern completion of related, overlapping representations (Marr, 1971; Leutgeb et al., 2007; Rolls, 2010; O'Reilly et al., 2011). Remarkably, the results further suggest that this tension between the representations can play an active part in the way that we perceive our episodic memories. In this case I infer these conclusions from differences across different individuals, but if the above explanation is correct, then these same processes should also account for differences within an individual, on different retrieval trials. While it was not possible to test this prediction on the current dataset, this would be an interesting avenue to explore in the future.

As well as the functional correlates described above, I also found that the same subjective differences could be predicted solely from differences in the anatomical size of CA3 across the group of healthy young adults. Again, this relationship was specific to CA3, and no other subfield showed a correlation with the behavioural variables. This result provides a striking demonstration that individual differences in something as seemingly subtle as subjective mnemonic perception can nevertheless be accurately predicted from measurable differences in anatomical structure. So what exactly does this tell us about the underlying neuronal mechanisms? It suggests that some function taking place within CA3 is enhanced by increased anatomical size,

and that this function itself is responsible for the changes in subjective perception. In other words, the relationship between CA3 size and subjective perception must be mediated by some functional process occurring within CA3. Given that I also found a strong correlation between Episodic Specificity within CA3 and the subjective awareness of commonalities, it seems likely that it is this function that mediates the relationship. In order to formally test this, I applied a mediation model, which confirmed the hypothesis, and demonstrated that CA3 size has a causal relationship with subjective distinctiveness through the informational content of CA3. This latter finding provides us with a clear model mapping the relationship between CA3 size, CA3 episodic information, and subjective mnemonic perception.

These results have important implications for our understanding of individual differences in episodic memory. They suggest that CA3 size can have a direct effect on core neural processes relating to episodic memory, which then lead to profound differences in the way that we perceive our memories. Several questions will need to be addressed if we wish to build on these findings. First, is there a clear relationship between this kind of subjective episodic distinctiveness and more objective measures of pattern separation, such as those used in Lacy et al. (2011)? If there is, can we find clear evidence that individual differences in both of these processes depend on structural differences in CA3? Second, why does an increase in CA3 size lead to more distinct episodic information? What exactly does this increased size mean in terms of the underlying neural architecture and processing? Finally, given the clear demonstration of hippocampal plasticity in taxi

drivers (Maguire et al., 2000; Woollett and Maguire, 2011), is it similarly possible to increase the size of CA3 through extensive training on e.g. a pattern separation task? Whatever the answer to these questions, it is becoming increasingly clear that gaining a deeper understanding of the role of the different hippocampal subfields in both animals and humans will be critical if we wish to fully elucidate the neural basis of episodic memory.

7 Chapter 7

Anticipating what is beyond the view: an fMRI study of boundary extension

Precis

The four experiments described so far focussed on the use of MVPA to investigate the nature of episodic representations within the hippocampus. A recent body of work suggests that vividly recalled episodic representations may depend on a process known as scene construction that takes place within the hippocampus. Scene construction is the process of mentally generating and maintaining a complex and coherent scene or event. This entails the retrieval of relevant components from modality-specific cortex, which are then channelled back into the hippocampus and bound into a coherent spatiotemporal representation. However, it is not currently clear how the process of scene construction contributes to, and interacts with, episodic representations within the hippocampus. In this final experiment I investigated the cognitive phenomenon of 'boundary extension' in order to better characterise the hippocampal role in scene construction, so that we might begin to understand the mechanisms by which this process contributes to episodic memory. Boundary extension is thought to depend on the automatic, implicit construction of scenes beyond the border of a given view. A recent study demonstrated that amnesic patients with selective bilateral hippocampal lesions showed reduced boundary extension, providing support for the idea that boundary extension depends on hippocampal scene construction processes. Here, I used a standard whole brain fMRI paradigm in order to investigate the neural correlates of boundary extension, thereby expanding our knowledge of the role of the hippocampus in automatic scene construction.

7.1 Introduction

In the natural world, what we see is always embedded within a wider context, and inherently part of an overarching sense of surrounding space. As such, we never perceive what is in front of our eyes in complete isolation, but instead each object is perceived as part of a visual scene, and each scene as one of an infinite set of related scenes that somehow form a continuous sense of place and space. In order to truly understand how we perceive the natural world, it is therefore crucial to understand how we process the world in terms of space and scenes. A central principle of visual perception is that visual input is necessarily limited and ambiguous. The brain overcomes this by making predictions about the likely content of the external world, extrapolating beyond the data that is directly available through the senses (Gregory, 1968, 1980; Friston, 2010). While scene perception is considered to take place at a relatively high level of the visual hierarchy (Epstein, 2008; Vann et al., 2009), there is evidence that even at this level the same principles of prediction and extrapolation apply. This is exemplified by a phenomenon known as 'boundary extension', whereby participants reliably remember seeing more of a scene than was present in the physical input, because they extrapolate beyond the physical borders of the original stimulus (Intraub and Richardson, 1989).

Boundary extension (BE) is a robust phenomenon that occurs across a variety of testing conditions including recognition, free recall, and even haptically (Intraub, 2004). It is apparent in all populations sampled including adults (Intraub and Richardson, 1989; Seamon et al., 2002),

children (Seamon et al., 2002; Candel et al., 2004), babies (Quinn and Intraub, 2007), and congenitally blind adults (Intraub, 2004). Importantly, BE only occurs in response to scenes, and not isolated objects (Intraub et al., 1998; Gottesman and Intraub, 2002). BE is a two-stage process; the first stage involves the active extrapolation of the scene beyond its physical boundaries, and is constructive in nature. This extrapolation occurs because when we initially encounter a scene, we are not limited to the direct sensory input entering the retina, but also have access to an automatically constructed and implicitly maintained representation of the scene. This constructed representation extends beyond the borders of the physical scene, and provides a global framework into which we can rapidly embed the salient elements from within the scene (Intraub, 2012). This process supports our experience of a continuous and coherent world, despite it being amassed from discontinuous sensory input, and is therefore highly adaptive under normal circumstances.

When the scene is no longer present, the extended scene content beyond the boundaries of the scene becomes incorporated into the internal representation of that scene. The second phase of BE occurs at retrieval, where the extrapolation beyond the original scene borders that occurred in the first phase is revealed by a subsequent memory error. Specifically, if presented with exactly the same scene a second time, we consistently judge the scene on this occasion to have less background, making it appear to be closer-up than the first scene. The fact that the studied view need only be absent for as little as 42ms for BE to be apparent (Intraub and Dickinson, 2008) underscores the online and spontaneous nature of this effect. The two

stages of BE are illustrated in Figure 33. The first stage, involving the active extrapolation of the scene beyond the boundaries, I hereafter refer to as the BE effect to differentiate it from the subsequent memory error, which I will refer to as the BE error.

Despite the insight into scene processing and prediction afforded by the BE effect, the neural underpinnings of this automatic extrapolation of scenes have not been well characterised. The only neuropsychological study of BE was conducted recently by Mullally et al. (2012), who found that selective bilateral damage to the hippocampi and concomitant amnesia were associated with attenuated BE compared to healthy control participants. This was consistently the case across a variety of different BE paradigms including both visual and haptic measures, demonstrating that this was a robust and cross-modal effect. This intriguing result suggests that the hippocampus may be critically involved in the BE effect, which is consistent with the known role of the hippocampus in spatial representation and scene construction (e.g. O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978; Burgess et al., 2002; Hassabis and Maguire, 2007, 2009; Hassabis et al., 2007a).

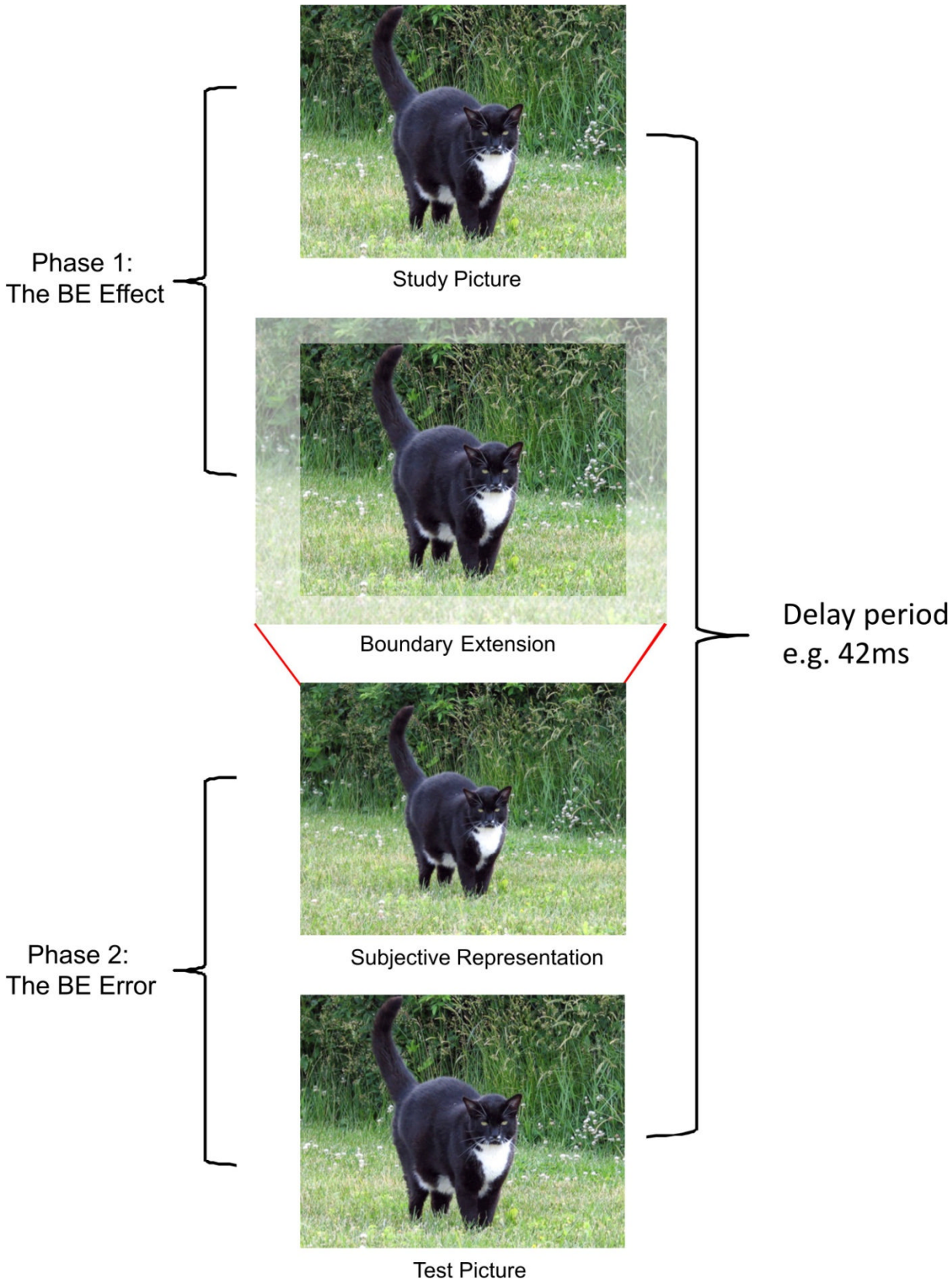

Figure 33. The two phases of boundary extension. *This figure demonstrates the processes that give rise to BE during a rapid serial visual presentation task. When a picture of a scene is presented for study, we automatically extrapolate beyond the physical edges of the scene (second panel). This active extension of the scene is the "BE effect". When the scene is no longer present, the extended content and context beyond the*

boundaries of the scene become incorporated into the subjective representation of that scene (third panel). Thus, in phase 2, when exactly the same picture is presented at test, we compare the now extended subjective representation to the actual picture, leading to a perception that the test picture is "closer" than the original study picture. This memory error is the "BE error".

Only one fMRI study has examined the neural correlates of BE, using a region-of-interest approach focused on two scene-relevant brain areas, the posterior parahippocampal cortex (PHC) and retrosplenial cortex (RSC) (Park et al., 2007). The aim of their study was not to investigate the activity relating to the initial active extension of a scene during the first presentation (the BE effect), but to investigate the neural adaptation effect on presentation of the second scene. Interestingly, they found that both the PHC and RSC demonstrated adaptation effects consistent with the subjective perception of the scenes rather than the physical reality. The results of this study suggest that these high-level scene-processing regions are sensitive to the output of BE at the BE error stage, as indexed by their sensitivity to subjective differences in perception. However, they do not allow us to draw any conclusions about the neural basis of the automatic extrapolation beyond the view of given scenes.

The aim of the current study was to investigate the active extrapolation of scenes occurring during the BE effect, and to shed light on the neural circuits involved in this kind of automatic scene construction. I used a modified version of a classic BE paradigm, known as the rapid serial visual presentation task, in which a picture of a scene was briefly presented onscreen, followed by a visual mask (Intraub et al., 1996; Intraub and Dickinson, 2008). After a brief interval (and unbeknownst to the participants)

exactly the same scene was presented for a second time, and the participant was required to decide whether the second scene appeared to be exactly the same as the first, or appeared to be closer or further away (see Figure 34). In order to investigate neural activity specifically related to the BE effect, I made use of the fact that the BE error does not occur on every single trial. This allowed me to compare trials where BE occurred to those where it did not. Specifically, I compared the activity elicited on the presentation of the first scene only on trials which subsequently led to a BE error to those first scene presentations which did not lead to a BE error. Regions involved in the active prediction and construction of extended scenes should show increased levels of activity on the trials where BE occurred compared to those where it did not. Based on the attenuated BE found in hippocampal amnesic patients, my central hypothesis was that the hippocampus would play a key role in this process (Mullally et al., 2012), with other high-level scene processing regions such as PHC and RSC potentially also making a contribution.

7.2 Methods

7.2.1 Participants

Thirty right-handed young adults (15 females) aged between 19 and 28 years of age (mean age 22.0 years; SD 2.88 years) participated in the experiment. All had normal or corrected-to-normal vision and gave informed written consent to participation in accordance with the local research ethics committee.

7.2.2 Procedure

During a pre-scan training period participants were instructed in the task requirements and were familiarised with the task during practice trials. I then collected fMRI data while the participants completed sixty trials of the task, presented in a randomised order. In a post-scan debriefing session, each participant demonstrated that they had fully understood the task and had made the intended responses.

7.2.3 Boundary extension task

The task was a modified version of a standard BE task, known as the rapid serial visual presentation (RSVP) task (Intraub et al., 1996; Intraub and Dickinson, 2008). At the start of each trial a central fixation cross appeared, indicating that the trial was starting. After 1s a scene picture was briefly presented in the centre of the screen for 250ms. This was then concealed with a dynamically changing visual noise mask which lasted for 200ms (Intraub and Dickinson, 2008). This was followed by a static visual noise mask presented for a variable period of 2, 3, or 4s. The length of this "jitter" was pseudo-randomised across trials. The purpose of this jittered period was to create separable neural signals for both the 1st and 2nd scene presentations (Dale, 1999). At the end of the jitter period a central fixation cross appeared for 1s, followed by a second scene picture presented in the same location. This was presented for 1s, after which it was joined by a set of options which appeared onscreen underneath the picture. Participants were provided with a scale of responses from 1-5, where 1 indicated that the second picture appeared to be "much closer-up" than the first picture, 2 that

it was “a little closer-up”, 3 that it was “the same” (the correct answer), 4 that it was “a little farther away”, and 5 that it was “much farther away” (see Figure 34). They were allowed up to 5s to make a response with a five-button scanner-compatible button-box using their right hand. Once they had made their response (or if they had failed to respond within 5s), a second set of options appeared, indicating that the participant had to provide a confidence rating regarding their decision. The rating was on a scale of 1-3, where 1 indicated that the participant was “not sure” of their response, 2 that they were “fairly sure”, and 3 that they were “very sure”; participants were allowed up to 4s to provide this rating. They were also given the option to press a button to indicate that they didn’t remember seeing the first picture at all. This was included given the rapid presentation of the first scene, and allowed for the fact that a participant may occasionally miss a scene due to lack of attention/long blinking at the crucial moment. Any trials on which a participant provided this response were discarded from the subsequent analysis, as were trials on which participant failed to provide a response to either of the ratings. Participants then had 2s to rest before the start of the next trial. On all trials, the second picture was identical to the first, although participants were unaware of this.

7.2.4 Behavioural analysis

I calculated a BE ratio score, which was the total number of trials judged to be “closer” (ratings of 1 and 2) minus the total number of trials judged to be “further” (ratings of 4 and 5), divided by the total number of trials. This provided a score from 1 to -1, where 1 indicates that every trial was judged to be closer, -1 indicates that every trials was judged to be further, and 0

indicates an even division between the two responses. In order to determine whether the group of participants as a whole displayed a significant BE effect, I compared the set of BE ratio scores to 0 using a one-way t-test.

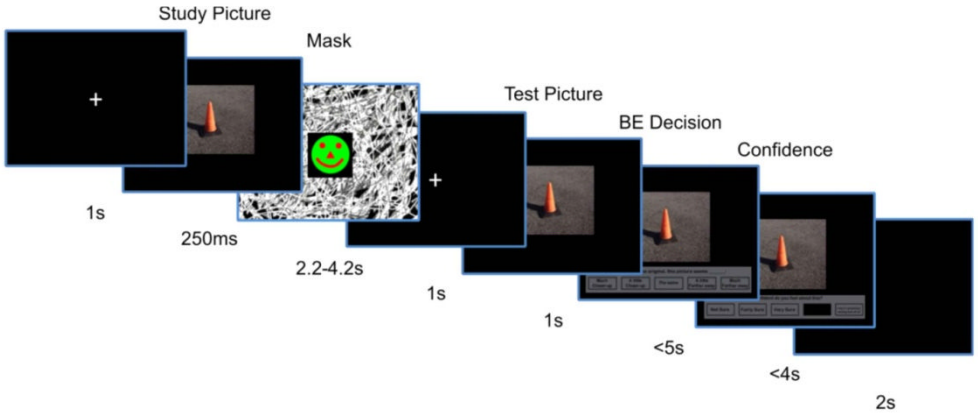

Figure 34. Example of a single experimental trial.

7.2.5 Anatomical regions of interest

My *a priori* hypothesis was that the hippocampus would be involved in the BE effect, and that the PHC and RSC might also show some involvement (particularly in the adaptation analysis – see later section). Each of these regions was manually defined on the normalized group average T1-weighted structural image (Figure 35), using the Duvernoy anatomical atlases for guidance (Duvernoy, 1999, 2005). These anatomical ROIs were used for planned small-volume correction in the whole-brain fMRI analyses (see later section). These same ROIs were also used for confirmatory analyses using MarsBar, and for the DCM analyses.

7.2.6 MRI acquisition

All MRI data were collected using a 3T Magnetom Allegra head-only MRI scanner (Siemens Medical Solutions) operated with the standard transmit-receive head coil. The whole-brain, standard resolution sequence (as described in Chapter 2) was used for all functional data acquisition. Field maps were acquired for distortion correction. T1-weighted MDEFT whole-brain structural scans were acquired for each participant after the main scanning session.

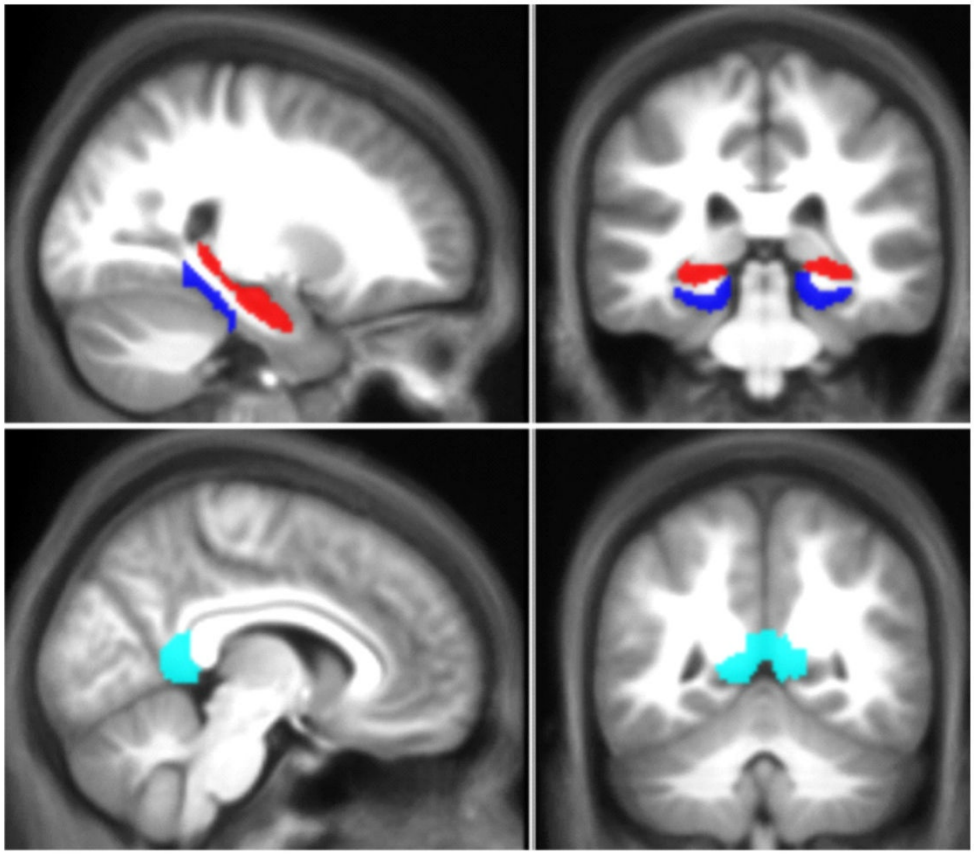

Figure 35. Anatomical regions of interest. *Following the results of Mullally et al. (2012), and Park et al. (2007), I defined three a priori anatomical regions of interest. These were the hippocampus (HC; displayed in red), posterior parahippocampal cortex (PHC; in blue), and the retrosplenial cortex (RSC; in cyan). Each of these regions was anatomically defined on the normalized group average T1-weighted structural image, using the Duvernoy anatomical atlases for guidance (Duvernoy, 1999; 2005). Here the three regions are displayed on the group average brain in the sagittal plane (leftmost images) and the coronal plane (rightmost images).*

7.2.7 Image pre-processing

All neuroimaging and statistical analyses were conducted using SPM8. The first six functional volumes were discarded to allow for T1 equilibration (Frackowiak et al., 2004).The remaining functional volumes were spatially realigned to the first image of the series, and distortion corrections were applied based on the field maps using the Unwarp routines in SPM (Andersson et al., 2001; Hutton et al., 2002). Each participant's structural scan was then co-registered to a mean image of their realigned, distortion-corrected functional scans. The structural images were then segmented into grey matter (GM), white matter (WM), and cerebral spinal fluid using the New Segment tool within SPM8. The DARTEL normalization process was then applied to the GM and WM segmented images, which iteratively warps the images into a common space using nonlinear registration (Ashburner, 2007). Using the output of this nonlinear warping process, all functional and structural images were normalized to MNI space using DARTEL's "Normalise to MNI" tool. The functional images were smoothed using a Gaussian kernel with full-width at half maximum of 8mm.

7.2.8 Neuroimaging analysis

Statistical analysis of the fMRI data was applied to the pre-processed data using a general linear model. The primary analysis involved a comparison of activity elicited by the first scene presentation on trials where boundary extension occurred to those first presentation trials where it did not. In order to do this, I used each participant's behavioural data in order to divide the trials into those where BE occurred (all trials where the second scene was

judged to be “closer” than the first - the BE condition), and those where it did not occur (all trials where the second scene was judged to be “the same” or “further” than the first - the Null condition). I used a stick function to model the onset of each first scene presentation, dividing the trials into two conditions based on the subsequent behavioural choice data, thus creating a BE regressor and a Null regressor. These stick functions were convolved with the canonical haemodynamic response function and its temporal derivative to create the two regressors of interest. I used a stick function to model the second scene presentations as well, also dividing them into BE and Null conditions, which were included as regressors of no interest. The BE decision period and confidence rating period were modelled as boxcar functions with variable length, depending on the participant-specific response times, and were included as regressors of no interest. Subject-specific movement parameters were also included as regressors of no interest.

In order to investigate the adaptation effects, a slightly altered version of this analysis was used. Instead of dividing the trials into BE and Null conditions, in this case I was interested in contrasting trials where the two scenes were perceived to the Same with those that were perceived to be Different (either “closer” or “further”). The trials were therefore divided into these two conditions for modelling both the first and second scene presentations. In all other respects the analysis was identical to the first. For each type of analysis, participant-specific parameter estimates (β values) were calculated at each voxel across the brain. The parameter estimates were then entered into a second level random effects analysis, whereby one-

sample t tests were applied to every voxel across the brain. Initial statistical thresholding was applied using a threshold of p = 0.001, uncorrected for multiple comparisons. Activations were considered to be statistically significant only if they survived family-wise error correction at either the peak or cluster level. Small volume corrections were applied to the *a priori* anatomical regions of interest.

7.2.9 ROI-based analyses

For each ROI analysis, the MarsBar toolbox[1] was used to fit the general linear models described above to the fMRI activation averaged across all voxels within a given region. Unsmoothed functional data were used as the input to these analyses in order to ensure anatomical specificity. MarsBar was also used to fit a finite impulse response model (Dale, 1999; Ollinger et al., 2001) to the data in order to probe the time-course of responses. Four time-windows of 2s each were modelled, time-locked to the onset of the first scene presentation.

7.2.10 Dynamic causal modelling

DCM is a Bayesian model comparison method which involves creating various plausible models of the task-dependent effective connectivity between pre-specified neural regions (Friston et al., 2003; Stephan et al., 2010). Once fitted, the evidence associated with each model can be compared in order to determine which is the most likely (or "winning")

[1] Brett M, Anton JL, Valabregue R, Poline JP (2002). Regions of interest analysis using an SPM toolbox. Abstract presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, Sendai, Japan. Toolbox available at http://marsbar.sourceforge.net/

model (see Chapter 2 for more details). I was interested in investigating the modulation of effective connectivity elicited by the presentation of the first scene on trials where BE occurred, and in order to do this I created a new, simplified design matrix for the DCM analysis, consisting of two regressors. The first modelled the onset of all first scene presentations, and the second modelled the first scene presentations on trials where BE occurred. I conducted two separate DCM analyses, in each case investigating the connectivity between two ROIs (hippocampus and PHC in one set of models, hippocampus and visual cortex in the second). I used DCM10 for these analyses, and in both cases the two ROIs were considered to have reciprocal average connections (the A matrix), with the visual input (the C matrix) stimulating the PHC in the first analysis and visual cortex in the second. For both analyses I created three different models based on altering the modulatory connections (the B matrix), allowing the modulation to affect the "backward" connection (from HC back to either PHC or visual cortex), the "forward" connection, or both directions (bidirectional). I conducted these analyses separately in both hemispheres, and used a random effects Bayesian model comparison method to determine which was the winning model (Stephan et al., 2009, 2010). This results in an "exceedance probability" estimate for each model, which describes how likely that model is compared with any other model. The model with the highest exceedance probability is considered to be the winning model.

7.3 Results

7.3.1 Behavioural results

The RSVP task produced a strong behavioural BE effect in the group of 30 participants, producing a mean average BE ratio score of 0.33 (SD = 0.22), which was highly significant ($p<0.00001$). Importantly, despite the strong overall BE effect, the proportion of trials on which the BE error occurred, averaged across subjects, was 48% (SD = 14%), thus providing a good division of the data into BE and Null trials for the main neuroimaging contrast.

7.3.2 Whole-brain fMRI results

I conducted a whole-brain fMRI analysis where I contrasted activity on first presentation trials where BE subsequently occurred to those where it did not. I focussed on activity evoked by the first scene presentation, as this is the time at which the active BE effect is proposed to take place. This analysis revealed significant activation within the right hippocampus (peak coordinate = 24, -39, 3; *Z* = 3.42; cluster size = 20), right PHC (peak coordinate = 21, -27, -18; *Z* = 3.71; cluster size = 46), and a significant activation extending across both left hippocampus and left PHC (peak coordinate = -26, -31, -14; *Z* = 3.45; cluster size = 35). Figure 36 displays each of these significant activations. No other activations reached significance, including within the RSC, indicating that this effect is specifically localised to the MTL.

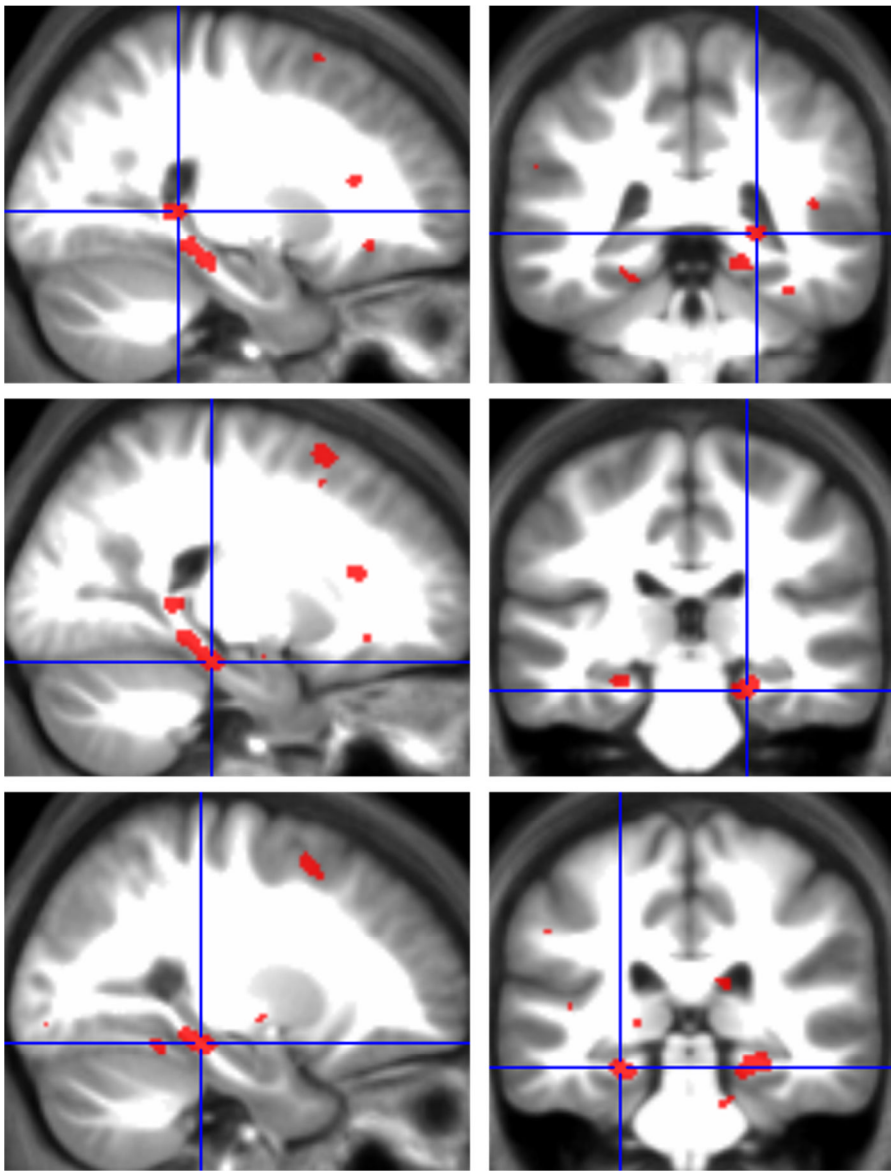

Figure 36. Neural correlates of the boundary extension effect. *Regions showing increased activation on trials where boundary extension occurred compared to those where it did not. This contrast was specifically focused on activity evoked by the first scene presentation. The three significant peaks are displayed separately, from top to bottom, in the sagittal plane on the left, and the coronal plane on the right, with the cross-hairs centred on the peak of the activation in each case. The top panel displays the activation in the posterior right hippocampus, while the middle panel displays the right PHC activation, and the bottom panel shows the activity in the left MTL spanning both hippocampus and PHC. For display purposes the activity is thresholded at p = 0.005 uncorrected. The results are displayed on the group average structural MRI scan.*

7.3.3 ROI analysis

I conducted a second analysis using MarsBar to extract the mean activation from each of the anatomically predefined ROIs (Figure 35), and this confirmed that bilateral hippocampus and bilateral PHC showed a significant increase in activation on trials on which BE occurred (one-tailed t-test statistics: left hippocampus $t = 1.98$, $p = 0.03$; right hippocampus $t = 2.08$, $p = 0.02$; left PHC $t = 2.72$, $p = 0.005$; right PHC $t = 3.49$, $p = 0.0008$). I also conducted a second analysis within MarsBar, focussing on the time-course of activation. The reason for this analysis is that, in order to assert that the activity reported here reflects the active extrapolation of scenes, it is important to establish that the significant MTL effect can truly be attributed to neural responses that are evoked by the first scene presentation. I therefore examined the time-course of activity within each region using a finite impulse response (FIR) analysis in MarsBar, which allowed me to look at the neural signal within specific time windows that are time-locked to the onset of the stimulus. I used four time windows of 2s each, time-locked to the onset of the first scene presentation on each trial.

Importantly, this analysis revealed a significant increase in activity on trials in which BE occurred as early as 2-4s following the first scene onset (one-tailed t-test statistics, averaged across hemisphere: hippocampus $t = 2.11$, $p = 0.02$; PHC $t= 1.94$, $p = 0.03$), indicating that this is an early response that is likely to occur soon after the stimulus onset (Figure 37). Given that the shortest delay between the onset of the first and second scene presentations was 3.45s (occurring on one third of the trials due to the jittered delay), I can conclude with some certainty that this effect during the 2-4s time-

window can only be attributed to a process occurring in response to the first scene.

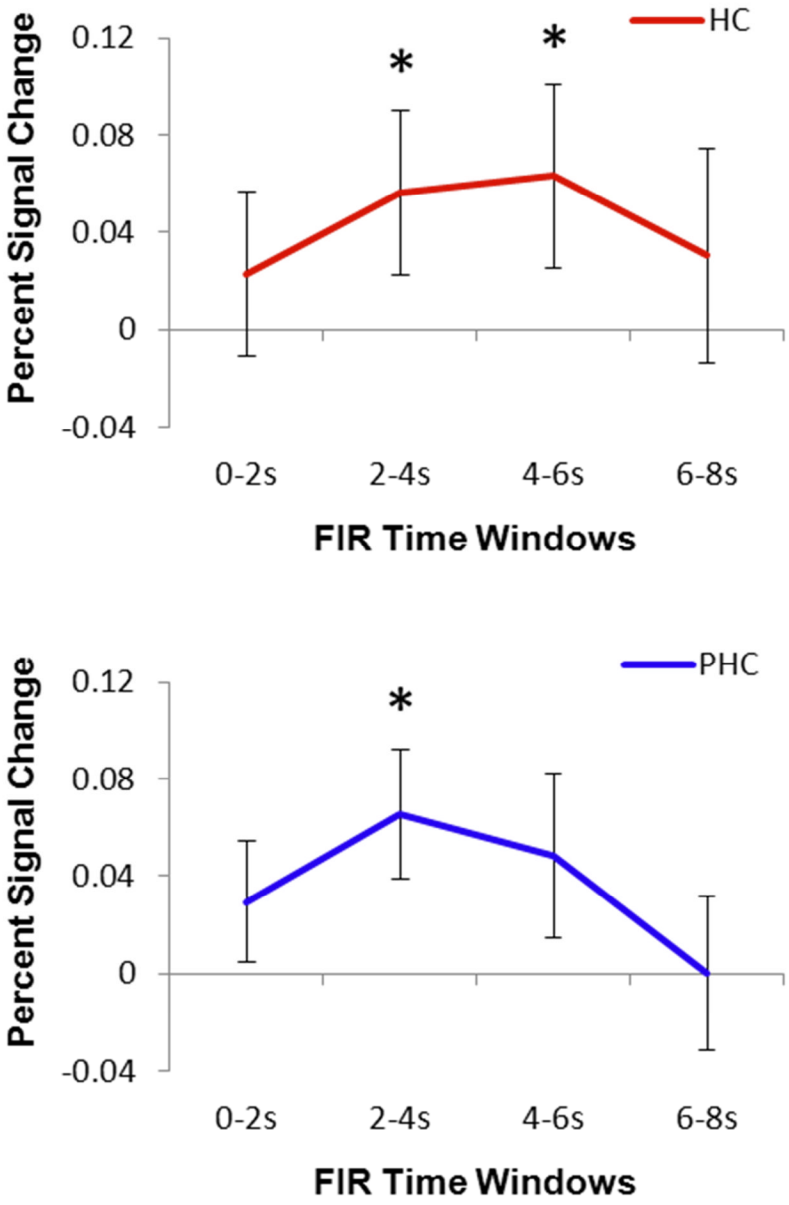

Figure 37. Time-course of the boundary extension effect. *A Finite Impulse Response (FIR) analysis was used to investigate the time-course of responses in the hippocampus (HC; top graph) and PHC (bottom graph). In each case the graph plots the increase in activity in each region on trials in which BE occurs compared to those where they do not. The different FIR time-windows are displayed on the x axis, and percent increase in signal change on the y axis. For both regions I found a significant increase in activation as early as 2-4s following the presentation of the first scene.*

7.3.4 Hippocampus– PHC connectivity

These results clearly demonstrate that both the hippocampus and PHC play an active role in BE. My original hypothesis was that the hippocampus should play a central role in the BE effect, because patients with damage localised to the hippocampus show reductions in BE (Mullally et al., 2012). I therefore wanted to tease apart the functional contributions of these two regions by investigating the neural dynamics occurring during the BE effect. If my hypothesis is correct, then I would expect the hippocampus to be driving the activity of the PHC. In order to assess the flow of information between these two regions, I used DCM, a Bayesian model comparison method in which different models of the neural dynamics are compared in order to find the most likely model of information flow in the brain (Friston et al., 2003).

For this analysis, I used the simplest approach possible, which involved investigating the connectivity between the two regions of interest: the hippocampus and the PHC (see Methods for more details). I conducted this analysis separately in both hemispheres, and used a random effects Bayesian model comparison method to determine which was the winning model (Stephan et al., 2009, 2010). The winning model was the backward modulation model, in which the hippocampus drove activity within the PHC, and this was the case for both hemispheres independently (exceedance probability for the backward model was 60% in the right, and 51% in the left hemisphere – see Figure 38). This result supports the original hypothesis, and suggests that the hippocampus is the driving force behind the BE effect, which then influences activity within the PHC.

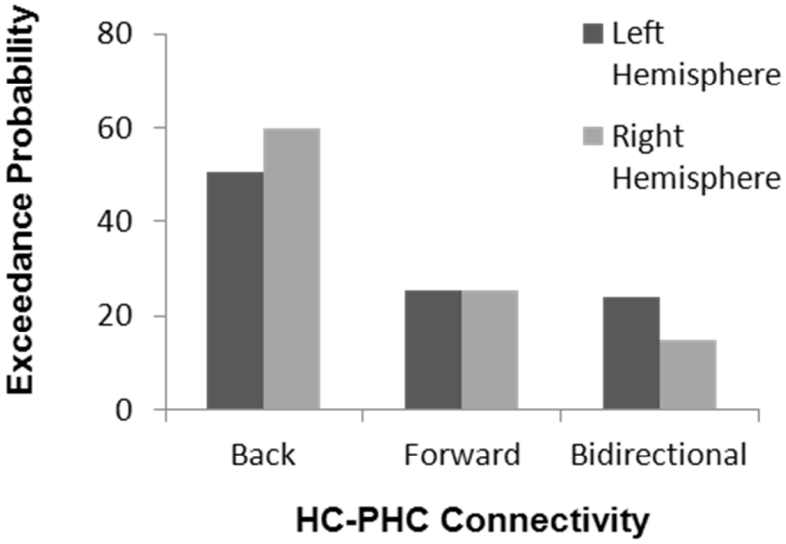

Figure 38. Modelling hippocampal-PHC connectivity during BE. *The results of the DCM model comparison analysis, displayed for both the left and right hemisphere. This plot displays the exceedance probability on the y axis, which describes how likely each model is compared to any other model. This is displayed for each of the three possible models: The "Back" model has the hippocampus influencing PHC, the "Forward" model has the PHC influencing the hippocampus, and the "Bidirectional" has a reciprocal modulation between the two regions. As hypothesised, the backward model was the winner in both hemispheres independently. This suggests that the hippocampus is the driving force behind the BE effect.*

7.3.5 fMRI adaptation

I have demonstrated that regions in the MTL are actively involved the extrapolation of spatial context during BE, with the hippocampus driving the PHC during this process. However, I was able to ask a further question with this dataset – is there any activity consistent with the subjective perception of scenes, similar to the results of Park et al. (2007)? In order to investigate this, I first searched for regions showing an overall effect of adaptation in response to scenes, regardless of the behavioural response. Interestingly, the only region in the entire brain to show an overall adaptation effect was a large cluster in early visual cortex (peak coordinate 6 85 3 – see Figure 39). I used MarsBar to probe the average activity in the

pre-defined ROIs, and this confirmed that none of the MTL regions displayed an overall adaptation effect in response to the scenes. In order to further investigate this adaptation effect within early visual cortex, I created an ROI using a contrast that was orthogonal to the subsequent adaptation analyses (all scenes presented on the first trial only compared to the implicit baseline).

Having defined this ROI, I next wanted to look for evidence of differential adaptation effects in line with subjective perception of the scenes. I therefore used MarsBar to extract the mean adaptation response on trials where participants perceived the second scene to be exactly the same as the first (no change in subjective perception) and those where the second scene was perceived to be different from the first (either closer or further away). Of the "different" trials, a group mean of 76% (sd = 17%) of the trials were perceived as "closer". If the early visual cortex displays responses that reflect the subjective perception of the scenes, we would expect this region to display less adaptation on trials where the scenes are perceived to be different compared to those which are perceived to be exactly the same. A direct comparison of the two adaptation responses revealed exactly this result (one-tailed t-test statistics: $t = 2.05$, $p = 0.03$), demonstrating that adaptation responses in early visual cortex track subjective perception even when there is never any physical change in the stimuli involved (Figure 39). A second analysis including only the "closer" trials in the "perceptually different" condition found the same significant difference in adaptation in early visual cortex (one-tailed t-test statistics: $t = 1.70$, $p = 0.05$).

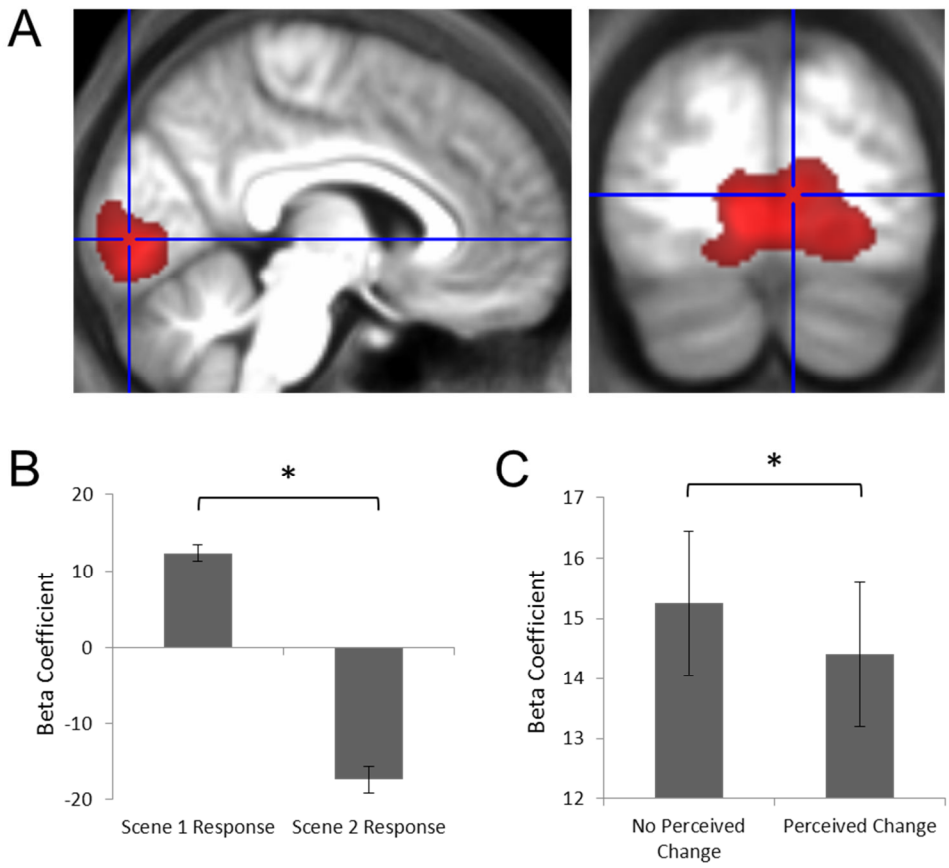

Figure 39. Adaptation effects in early visual cortex reflect changes in subjective perception. *(A) A whole-brain analysis investigating fMRI adaptation effects between the first and second presentation of the scenes. The only significant activation was in early visual cortex, here displayed at a family wise error corrected threshold of p = 0.05 on the group average structural MRI scan. The crosshair is centred on the peak of the activation. (B) The average response within this visual region to the first and second scene presentations, with standard error bars. This plot demonstrates that visual cortex shows a robust adaptation effect to repeated scene presentations. The y axis displays the parameter estimates from the general linear model. (C) The magnitude of the adaptation effect (i.e. the amount of attenuation between first and second scene presentation) for the two conditions of interest, with standard error bars. When participants perceive a change between the first and second scene presentation (e.g. when it appears to be "closer") there is a significant reduction in the magnitude of adaptation compared to trials where participants perceive no change between the two scenes. This is despite the fact that the two scenes are always physically identical. The y axis displays the contrast between the parameter estimates for the 1st and 2nd scenes.*

Although no MTL regions displayed any evidence of an overall scene adaptation effect, I nevertheless investigated whether the PHC and RSC might display a *differential* adaptation effect in line with the results of Park et al. (2007). I found that both regions showed differential adaptation in line with the subjective perception of the scenes, so that they showed less adaptation for scenes perceived to be different (one-tailed t-test statistics, averaged across hemisphere: PHC $t = 1.81$, $p = 0.04$; RSC $t = 1.7$, $p = 0.05$). Thus, although these regions did not show a global adaptation effect in response to repeated scenes, they nevertheless showed the expected pattern of differential adaptation. These results, therefore, are broadly consistent with the results of Park et al. (2007), and suggest that both the PHC and RSC display activity that tracks the subjective perception of scenes. This was not the case with the hippocampus, which did not display a significant difference in adaptation ($t = 1.43$, $p = 0.08$).

7.3.6 Top-down modulation of visual cortex – DCM results

Overall the results thus far suggest that the MTL, and particularly the hippocampus, is involved in the rapid, automatic extrapolation of scenes beyond the edges of the given view. For visual cortex to show this kind of differential adaptation response to trials perceived to be the same and trials perceived to be closer (i.e. trials where BE occurred), the subjective scene representations, including the extended boundaries, must be made available to this region before the onset of the second scene via some top-down connection.

This suggests that the early visual cortex may be actively influenced by the hippocampus following the first scene presentation, during (or shortly after) the BE effect itself. In order to investigate this possibility, I applied a DCM analysis to the neural dynamics of the hippocampus and early visual cortex during the presentation of the first scene. If the hippocampus is actively involved in updating the visual representations of scenes to include the extended boundaries in line with subjective perception, then we would expect to find evidence for modulation of visual cortex by the hippocampus on those trials where BE occurred. I compared this model to two alternative models (modulation of hippocampus activity by visual cortex, and bidirectional modulation), and found strong evidence in favour of backward modulation of visual cortex by the hippocampus. I analysed each hemisphere separately, and found robustly consistent results across both hemispheres (displayed in Figure 40). These results therefore confirm my hypothesis that activity in early visual cortex is modulated by the hippocampus when BE occurs, and that this modulation occurs at the time of, or shortly after the active extrapolation beyond the borders of a scene.

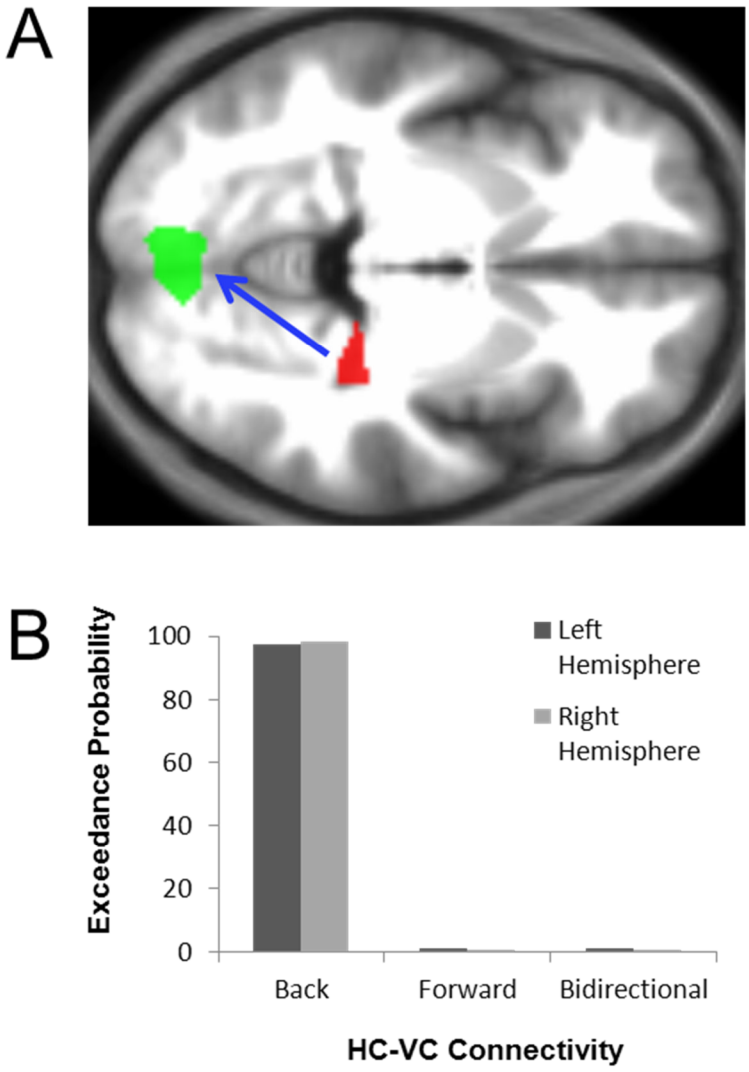

Figure 40. Modelling hippocampal-visual cortex connectivity *(A) The hypothesised flow of information, with activity in early visual cortex being actively modulated by the hippocampus during the BE effect. The hippocampus is displayed in red and visual cortex in green on an axial slice from the group average structural MRI scan. (B) The results of the DCM model comparison analysis, displayed for both the left and right hemisphere. This plot displays the exceedance probability on the y axis, which describes how likely each model is compared to any other model. This is displayed for each of the three possible models: The "Back" model has the hippocampus influencing visual cortex, the "Forward" model has the visual cortex influencing the hippocampus, and the "Bidirectional" has a reciprocal modulation between the two regions. As hypothesised, the backward model was the clear winner, demonstrating an exceedance probability of more than 97% independently across both hemispheres. This suggests that the hippocampus actively updates the scene representations within early visual cortex following boundary extension.*

7.4 Discussion

This study generated several new findings. First, I found that both the hippocampus and PHC were active during the automatic extrapolation of scenes beyond the given view, and that this process appeared to occur online, while a scene was physically present. An analysis of the flow of information between these two regions suggested that this process was driven primarily by the hippocampus, which then influenced activity within PHC.

Previous studies have linked the hippocampus to the perception of complex scenes. For instance, patients with selective bilateral hippocampal damage have been found to show deficits in the ability to discriminate complex scenes, but not complex objects or faces (Lee et al., 2005a, 2005b), demonstrating that the hippocampus plays an important role in the perception of scenes, but not single objects. Similarly, an fMRI study of a scene discrimination task found activation in the posterior hippocampus and PHC, demonstrating that both of these regions are involved in scene processing in healthy individuals (Lee et al., 2008). This conclusion was supported by the results of a more recent MVPA study which found that the hippocampus maintains representations of individual complex natural scenes (Bonnici et al., 2012). Thus, there is growing evidence to suggest that the hippocampus plays a role in the online perception of complex scenes.

A second set of studies demonstrates that the specific role of the hippocampus may go beyond passive scene processing, and extend to the active construction of scenes (Hassabis et al., 2007a, 2007b). Hassabis et al. (2007a)

found that patients with selective bilateral hippocampal damage and amnesia were unable to imagine novel spatially coherent scenes (see also Andelman et al., 2010; Race et al., 2011), which led to the proposal that the hippocampus may be involved in the construction of complex spatial contexts or scenes into which sensory objects and events are bound (Hassabis and Maguire, 2007, 2009). Most recently, Mullally et al. (2012) found that patients with hippocampal lesions showed reduced levels of BE, which depends on scene construction.

The results presented here provide further evidence that the hippocampus is actively involved in the BE effect in healthy adults as well as amnesic patients. Importantly, my findings also demonstrate that the hippocampus is involved in boundary extension at the time of the first exposure to a scene, confirming that its involvement is in the initial phase of active anticipation of what is beyond the view rather than any subsequent memory-related effect.

Natural scenes are highly complex stimuli composed of multiple elements, and one important question we might ask is exactly what aspect of scene representation and anticipation is supported by the hippocampus. A wealth of evidence from the animal literature demonstrates that the hippocampus plays an essential role in the representation of the spatial environment (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978; Andersen et al., 2006). Evidence has also accumulated suggesting that the hippocampus plays a similarly critical role in spatial representation in humans as well. For instance, trainee London taxi drivers display enlargement of the posterior

hippocampus over the years of training it takes to gain “the knowledge” (comprehensive knowledge of the layout of London’s complex street network, landmarks, and routes within London), providing evidence that the hippocampus is critically involved in the representation of spatial information (Maguire et al., 2000; Woollett and Maguire, 2011). Similarly, fMRI studies have demonstrated that the hippocampus is important for route planning during navigation (e.g. Hartley et al., 2003; Spiers and Maguire, 2006; Viard et al., 2011) and representing locations within an environment (e.g. Hassabis et al., 2009). Given the clear importance of the hippocampus to spatial representation, it may be that the role of the hippocampus in scene processing and boundary extension is also spatial.

Evidence for this comes from two sources. First, Hassabis et al. (2007a) probed the nature of the scene construction deficits in their group of hippocampal patients, and found that the impairment was most pronounced in the spatial coherence of imagined scenes. In other words, the patients’ attempts at scenes were spatially fragmented, suggesting that the deficit may be primarily spatial in nature. Second, there is evidence that the hippocampal role in boundary extension itself is also likely to be spatial. Mullally et al. (2012) included a control task whereby they explicitly tested the ability of hippocampal amnesic patients to extrapolate beyond the edges of a given scene. In this “scene probe” task, they found that the patients were able to accurately describe the given scene that was directly in front of them. Then, when asked to imagine taking a step back from the current position and describe what might then come into view, the patients' performance was comparable to the control participants. They were able to list contextually rele-

vant items in the extended scene, associated them with one another, and could relate them to the context. However, they were unable to visualize these additional items within a spatially coherent extended scene. Thus, while the capacity for basic semantic, conceptual, associative, and relational processing was intact for an imagined scene beyond the edges of a picture, the spatial coherence of the extended scene was impaired. This therefore suggests that the role of the hippocampus in boundary extension is likely to be spatial in nature. The hippocampus may provide the spatial framework of a scene into which the scene elements can be bound, such that when a scene picture is presented, the hippocampus anticipates the space beyond the edges. The extended space is then filled in with additional scene content, leading to the BE effect. The current results demonstrate that this anticipatory process occurs rapidly and automatically during online perception of a scene.

A second important finding from my study comes from the adaptation analysis, which demonstrated that the activity of several key regions reflected the subjective perception of scenes. Consistent with the results of Park et al. (2007), I found that two high-level scene-processing regions, the RSC and PHC, both showed activity profiles that mapped onto subjective perception. This result suggests that these regions do not simply contain veridical representations of the physically presented scenes, but are actively updated to include information about extrapolated scene content beyond the boundaries of the physical scenes.

Intriguingly, I found that early visual cortex also displayed differential fMRI adaptation effects that reflected the subjective perception of the scenes. Specifically, visual cortex showed greater adaptation when no change was perceived between two scene presentations, compared to those trials where the scenes appear to be closer (consistent with the BE error). Importantly, the two scene presentations were always identical, so this effect cannot be attributed to any physical changes in the stimuli but can only be due to a change in subjective perception driven by top down process. This latter result is consistent with a variety of studies which have shown that activity as early as V1 can reflect changes in subjective perception (Tong, 2003; Kamitani and Tong, 2005; Murray et al., 2006; Sperandio et al., 2012), and demonstrates that this can also be the case with the perception of complex scenes.

The third finding from this study revealed that activity within early visual cortex was modulated by a top-down connection from the hippocampus at the time of the BE effect. This modulation suggests that the scene representation within visual cortex is actively updated by a top-down connection from the hippocampus to represent the extended scene. This updated (subjective) representation then leads to the subsequent differential adaptation effect. It is worth nothing that Park et al. (2007) also looked for similar adaptation results within retinotopic cortex and failed to find any evidence for such an effect. I believe that the differences between mine and their findings in this regard are likely to arise from differences in the study designs. Specifically, Park et al. (2007) used an implicit task in which inferences were made on the basis of different conditions which, on average,

produced different degrees of the BE effect. By contrast, I recorded explicit trial-by-trial behavioural choice data, which allowed me to directly compare trials which individuals perceived as the same to those where BE occurred. My approach is likely to have provided substantially greater power to detect activity relating to subjective perception of scenes within early visual cortex.

Put together, my findings offer a new insight into the neural basis of scene processing. They suggest a model of scene processing whereby the hippocampus is actively involved in the automatic anticipation of 'unseen' aspects of scenes which are then channelled backwards through the processing hierarchy via PHC and as far as early visual cortex in order to provide predictions about the likely appearance of the world beyond the current view. This subsequently leads to a differential adaptation effect within early visual cortex which is driven solely by a subjective difference in appearance due to the extended boundaries. The fact that the information about the extended scene is automatically and rapidly channelled as far back as early visual cortex suggests that this anticipation of scenes is a pervasive and important process in our online perception. While this may seem surprising, it is entirely consistent with theories positing the central role of prediction to our activities in the natural world (Gregory, 1968, 1980; Friston, 2010). Given that our world is inherently spatial in nature, the prediction of spatial context may form a critical part of these predictions, and this may be the fundamental contribution of the hippocampus to scene construction and BE, although further studies are required to establish this directly.

8 Chapter 8

General Discussion

Despite a long history of research into functions of the hippocampus, we still have little concrete knowledge regarding the neuronal representation of individual episodic memories. This is largely because current methods of studying brain function in humans do not permit the examination of individual episodic memories in terms of the neuronal population activity that underpins them (Gelbard-Sagiv et al., 2008; Hassabis et al., 2009), with the further complication that true episodic memory can only be studied with certainty in humans (Tulving, 2002; Suddendorf and Busby, 2003). Thus, while theoretical models of hippocampal function abound (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011), empirical evidence pertaining to episodic representations currently lags behind.

Given this background, the central aim of my thesis was to explore the nature of the information represented in the human hippocampus, with a particular focus on episodic representations. To do this, I conducted five fMRI experiments, each designed to address a specific experimental question related to the overarching theme of hippocampal episodic representations. Each study therefore informed a different aspect of the overall question "what information is represented in the human hippocampus?". Having already discussed each set of results in the relevant chapter, I now pose the experimental questions again, and on this occasion consider some of the broader issues raised by my results. Then, in the concluding section, I discuss outstanding questions and suggest directions for future research into episodic representations.

8.1 Can we detect individual episodic memory representations in the human hippocampus?

As mentioned above, and described in detail in Chapter 1, we currently have very little concrete evidence about how an individual episodic memory is represented in the human brain. In Experiment 1 (Chapter 3), I applied MVPA to the analysis of high-resolution fMRI data in order to investigate the representation of individual episodic memories in the human hippocampus. This study provided the first evidence that it is possible to detect information about individual episodes solely from the pattern of activation across voxels in the hippocampus during episodic recall. This showed that it is indeed possible to measure information about specific episodic memories, and thereby investigate the neural representation of individual episodic memory traces. Furthermore, I found that the hippocampus contained significantly more episodic information than either the entorhinal cortex or posterior parahippocampal cortex. Consistent with the functional and anatomical hierarchy of the MTL (Rolls, 2010; O'Reilly et al., 2011), this suggests that the episodic representations are more distinct within the hippocampus itself than cortical MTL regions. Finally, within the hippocampus itself, episodic information was not evenly distributed. Instead, it was biased towards three regions, in the right posterior and bilateral anterior hippocampus.

From these results, I concluded that it is possible to detect information relating to specific neuronal populations within the hippocampus, each of which codes for an individual episode. However, is this conclusion valid? Given that we are measuring BOLD activation at a spatial scale that is orders of magnitude larger than the individual neurons themselves, what can we truly infer from such results? This is an important question, as ultimately we wish to use the results of MVPA to make inferences about representations at the level of neuronal population activity.

There is currently some debate as to the nature of the underlying signal in MVPA analysis of fMRI data (Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010; Op de Beeck, 2010a, 2010b). Some argue that MVPA is able to detect information about neuronal population activity that is below the spatial scale of the voxels themselves (Haynes and Rees, 2005, 2006; Kamitani and Tong, 2005; Kamitani and Sawahata, 2010). This argument is most frequently proposed for decoding analyses of grating orientation from patterns of voxels in primary visual cortex. Orientation-selective neurons cluster together into discrete cortical columns, and the proposal is that random variation can lead to some voxels sampling more columns of one orientation than another. This would lead to the result that the activity in this voxel as a whole will be biased towards that orientation. If the same principle applies across a set of voxels, this may provide enough information for MVPA to reliably decode grating orientation. The columnar organisation of primary visual cortex provides a clear, concrete example of this hypothesis, but the same basic principle could apply to MVPA applied to other regions as well. For example, neuronal populations coding for a

particular memory might be randomly distributed across the hippocampus, but biased in some voxels over others, thereby providing information at the level of voxel patterns of activity. On the other hand, some have argued that the information may in fact be based on weak, larger-scale topographical organisation, rather than random biasing of small-scale neuronal populations (Op de Beeck, 2010a, 2010b). This debate is further complicated by the influence of the neural vasculature. The fMRI BOLD signal is an indirect measure of neural activity, and the precise location, scale, and amplitude of this signal can be biased by local vasculature (Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010). At the scale of univariate fMRI, this bias is generally not relevant, however it does cloud the above argument relating to neural information at the sub-voxel scale. Thus, MVPA analyses do not easily allow one to infer on the precise type or scale of the underlying neuronal activity.

Fortunately, it is not essential for our purpose that we can infer the precise form of the underlying neural signals. Rather, what is important is that we can infer the presence of some form of specific neural information relating to each individual episodic memory. The presence of a significant MVPA result tells us that there must be some form of neuronal activity causing reliable, detectable signals in the multi-voxel patterns of activity. Thus, while we cannot directly infer the type of neuronal activity, or the precise spatial scale of that activity, we can infer that this activity contains information about specific cognitive states (i.e. individual episodic memories), which is the critical inference. Overall, therefore, I can conclude with some certainty that MVPA allowed me to detect individual episodic

representations within the human hippocampus.

8.2 How do episodic memory representations change over time?

While it is well-established that the hippocampus is critical for the representation of episodic and autobiographical memories for at least a certain period of time, it is currently a matter of debate as to whether more remote memories are also dependent on the hippocampus (Squire, 1992; Nadel and Moscovitch, 1997; Squire et al., 2004; Moscovitch et al., 2005; Winocur et al., 2010; Winocur and Moscovitch, 2011). In the second experiment (Chapter 4), I applied an MVPA approach in order to compare the neural representation of recent autobiographical memories from two weeks previously, and remote memories that were more than a decade old. In so doing, I was able to provide evidence to inform this long-standing debate, and shed new light on the changes that episodic representations may undergo with the passage of time.

This study produced several novel results. First, information about individual autobiographical memories was present within the hippocampus regardless of age. Thus, whether the memories were two weeks old or more than a decade old, distinct, reproducible patterns of voxel activity were detectable within the hippocampus during vivid recall. This result is clearly not compatible with the theory that episodic memories are consolidated out of the hippocampus into neocortex, and therefore speaks against the

standard model of consolidation (SMC). At the same time however, I found that distributed cortical regions such as the ventromedial prefrontal cortex (vmPFC) and temporal pole (TP) showed a clear increase in the strength of memory representation for remote compared to recent memories (the recent memories were also detectable within these regions but decoding accuracy was significantly less). This result is consistent with some form of consolidation process occurring within neocortex. Finally, I found that the posterior hippocampus also displayed an increase in the strength of memory representation for remote compared to recent memories. I have discussed possible interpretations of these results in some detail in Chapter 4, and I will not reprise this in detail here. Instead I will extend the discussion further by elaborating on one specific conceptual issue surrounding these results, and fMRI studies of remote episodic memory in general.

It could be argued that the presence of distinct representations of remote memories within the hippocampus does not in itself tell us that these representations are functionally relevant or necessary for the retrieval of those memories. For example, a proponent of the SMC could argue that the memory itself has been consolidated to the neocortical regions, and that residual activity within the hippocampus itself is merely epiphenomenal. This case is easier to make against the results of univariate fMRI studies, where the signal is not specific to individual episodic representations. With MVPA, this argument is more difficult to make, as I have demonstrated the presence of distinct episodic representations of remote memories. Nevertheless, one could take the extreme view that even these distinct memory traces have become functionally irrelevant as the cortical regions

have taken over the functionally relevant representation of the memories.

However, here I argue that it is not reasonable to suggest that selective neuronal activity that can be mapped onto a specific cognitive process or representation is functionally meaningless. Such a position is essentially arguing that an experimentally specific increase in activation (or pattern of activation) in a brain region is contributing nothing to a cognitive state. This view is against one of the central tenets of neuroscience – that selective neural activity that can be reliably linked to a cognitive state must make some functional contribution to that cognitive state, even if that contribution is subtle (e.g. Logothetis, 2008). In my view, therefore, this SMC-type argument is untenable.

If we accept this position, then given the clear presence of remote memory representations in the hippocampus, we must accept that these representations have some functional relevance. I believe, therefore, that it is time we moved away from arguing about whether or not the hippocampus is involved in the representation of remote memories, and instead ask the more pertinent question - what function is this activity contributing to episodic recall? Given the mixed results of the patient studies, where some patients show impaired remote memory retrieval, and others do not (for a review, see Winocur and Moscovitch, 2011), it may be that some form of episodic retrieval is possible after selective hippocampal damage (although note that this depends critically on precisely how episodic memory is defined and assessed – Winocur and Moscovitch, 2011). The critical question, therefore, is to determine exactly what functional contribution the

hippocampus makes to remote episodic recall. Various competing theories make different predictions regarding the likely contribution. For instance, MTT argues that the hippocampus is always required for richly detailed, vivid episodic retrieval, regardless of age (Moscovitch et al., 2005; Winocur and Moscovitch, 2011). Scene construction theory (SCT), on the other hand, proposes that the hippocampus facilitates the spatial framework into which the elements of a remote memory are bound and re-constructed (Hassabis and Maguire, 2007, 2009). Thus, according to SCT it is the spatial coherence of a memory which is critically determined by the hippocampus, rather than vividness *per se*. Another potentially important variable is the precise location of damage within the hippocampus. In this study I demonstrated increased strength of episodic representation within posterior hippocampus for remote compared to recent memories. This suggests that lesions of the anterior and posterior hippocampus may have dissociable effects on recent and remote memory retrieval, with posterior hippocampus damage leading to greater remote memory deficits. At present it is not clear precisely which mechanism can explain the functional role of the hippocampus in remote memory, but I believe that appropriately designed fMRI studies will play a crucial part in better defining this role.

To return to the specific results of Experiment 2, as described above, I found that all regions studied, including the hippocampus, contained representations of both the recent and the remote memories, indicating that the episodic memories do not become fully independent of the hippocampus. I also found that neocortical regions such as vmPFC and TP, along with the posterior hippocampus, displayed increased levels of episodic representation

with the remote compared to the recent memories. This set of results as a whole provides novel evidence about the transformation of episodic memory traces over time, and I argue that the set of results is consistent with two important conclusions. First, that episodic memories do not become fully independent of the hippocampus, even after ten years. Second, that the increased representation of remote episodic memories in posterior hippocampus is consistent with a scene construction process, which is required for both recent and remote memories, but more so for remote memories. This latter conclusion is currently speculative, but does nevertheless fit the available data better than competing theories (for a more thorough discussion, see Chapter 4).

8.3 What is the nature of episodic memory representations in the hippocampus?

The first two experiments demonstrated that MVPA analyses could detect distinct episodic representations within the hippocampus. However, in both studies, the memories that were examined differed along a variety of dimensions including spatial location, their content, and the people involved. It is therefore possible that the MVPA analyses could be detecting any one of these sources of information (or a combination of them) in order to decode the memories. Thus, it was not possible to determine exactly what information was being decoded in these previous studies, which limits the ability to draw inferences about the nature of the episodic representations. In the third experiment (Chapter 5) I created a controlled set of episodes in

order to try and determine more precisely the nature of episodic representations in the hippocampus.

Theories of hippocampal function propose that the hippocampus represents episodic memories as distinct memory traces, even when those episodes overlap with one another in terms of their content (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). However, this hypothesis has not been directly tested with complex episodic memories. In this experiment, I used green-screen technology in order to superimpose two short movie clips against two spatial backdrops, creating a set of four perfectly controlled episodes. Due to the overlapping spatial context and event content, it was not possible to distinguish any individual episodes solely on the basis of these constituent elements. This could only be accomplished by representing each episode as a distinct episode, over and above the constituent elements.

I found that it was possible to decode vividly recalled memories of these episodes from the hippocampus (but not from other MTL regions) despite the high degree of overlap. This result can only be due to the presence of distinct, conjunctive representations of each of the four episodes within the hippocampus. This result therefore supports the hypothesis that the hippocampus represents episodes as distinct memory traces, even in the presence of overlapping elements (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). The overlapping episodes allowed me to ask additional questions about episodic representation. Specifically, I was able to test for the presence of generalised

spatial context information that was shared across different episodes. Interestingly, I found evidence for these representations within the hippocampus as well, but not within the other MTL regions. I found no such evidence for generalised event content information.

Put together, this set of results provides potentially important information about the nature of episodic representations within the hippocampus. First and foremost, it confirms the long-standing hypothesis that the hippocampus contains distinct, conjunctive representations of different episodes, even in the presence of overlapping elements (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011). Second, in addition to these unique episodic representations, the hippocampus also maintains general representations of spatial context, which may be shared across different episodes. This latter result is consistent with the well-established role of the hippocampus in spatial representation and spatial navigation (e.g. O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978; Burgess et al., 2002; Hassabis et al., 2007a, 2009), and also suggests that the hippocampus is able to maintain multiple types of representation during episodic recall. As I discuss in the next section, these results have implications for the proposed computational role of the hippocampus, and in particular the contribution of the individual subfields of the hippocampus.

8.4 How do hippocampal subfields contribute to episodic memory representations?

In the first three experiments I investigated episodic representation within the hippocampus as a whole. However, the hippocampus is made up of several constituent subfields, each with its own distinct cytoarchitecture and pattern of connectivity (Lorente De No, 1933, 1934; Duvernoy, 2005). The connectivity between these subfields results in a unique circuit, and one influential theory argues that the computations conducted within each separate subfield, combined with this distinctive circuit connecting the subfields, form the basis of the hippocampus' role in episodic memory (Marr, 1971; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 2010; O'Reilly et al., 2011).

There are two key computations that are proposed to account for many of the mnemonic abilities of the hippocampus. The first is pattern separation, whereby the DG orthogonalizes incoming information, which can then be stored as distinct representations within region CA3. This process is particularly important for episodic memories, which are expected to contain much overlap with one another due to the necessarily limited scope of day-to-day experience. The second critical computation is pattern completion, where a partial cue can trigger the retrieval of an entire memory representation within CA3, which will then trigger the memory representation stored within CA1, from where the entire distributed memory

representation will be reactivated.

There is a growing body of evidence from the rodent literature suggesting that neural computations such as pattern separation and completion may take place within the subfields of the hippocampus (Lee et al., 2004; Leutgeb et al., 2004, 2007; Vazdarjanova and Guzowski, 2004; Wills et al., 2005). More recently, fMRI studies have produced evidence consistent with pattern separation processes occurring within human CA3/DG in response to pictures of objects with graded levels of similarity (Bakker et al., 2008; Lacy et al., 2011). While these studies provide evidence in support of the existence of these computations, no empirical evidence exists to support a direct link between this theoretical account and episodic memory. Thus, in Experiment 4 (Chapter 6) I investigated the episodic representations present within the individual subfields of the hippocampus in order to test the computational models of episodic memory in the human hippocampus.

In order to do this I turned to the overlapping episode dataset collected for Experiment 3. The sub-millimetre, high-resolution structural images allowed me to delineate four major subfields of the hippocampus (CA1, CA3, DG, subiculum) in each subject, thereby permitting a comparison of episodic information between the subfields. Because the episodes in this dataset were precisely controlled yet overlapping, I could assess two types of representation within the subfields. First, I looked for evidence of unique, conjunctive episodic representations for each of the four episodes. As predicted by the theoretical accounts, I found that regions CA3 and CA1 contained more distinct episodic information than the other two subfields.

This suggests that these subfields in particular are important for the retrieval of unique episodic memory traces. The overlapping nature of the episodes also allowed me to look for evidence of episodic pattern completion within each of the subfields. I found that of all the subfields, only CA3 showed any evidence for activation of the overlapping episodes during episodic retrieval. Again, this result is entirely consistent with the computational account of hippocampal function. Put together, these results offer strong support for the theory that neural computations taking place within the hippocampus may explain important aspects of complex episodic memory, such as the ability to represent overlapping episodes as distinct memories.

In addition to the results described above, I also found that individual differences in the perceived distinctiveness of the four overlapping episodes could be predicted on the basis of both the amount of distinct episodic information within region CA3, and also the anatomical size of CA3. None of the other subfields displayed any such relationship, demonstrating that the results were specific to CA3. A formal mediation analysis demonstrated that the differences in mnemonic perception were caused by the size of CA3, mediated by the episodic information within CA3. Not only do these results lend further credence to the conclusion that CA3 is particularly important for the neural computations described above, but they also provide new information about how individual differences in how we perceive our own memories can be influenced by the neural processing and anatomy of specific hippocampal subfields.

Overall, my results provide the first empirical link between the computational models of hippocampal function and complex episodic memories. This crucial link thereby validates the use of these models for providing an explanatory framework for the neural basis of episodic memory. This may now pave the way for putting the study of human episodic memory onto a more quantitative and rigorous theoretical footing.

8.5 What is the role of the hippocampus in boundary extension?

Episodic representations in the hippocampus are thought to depend on a process known as scene construction (Hassabis and Maguire, 2007, 2009). Scene construction involves the mental generation of a complex and spatially coherent scene or event, which is proposed to be a critical component of rich, vivid episodic recall. Several studies have demonstrated that patients with bilateral hippocampal lesions show impairments in the ability to construct novel scenes (e.g. Hassabis et al., 2007a; Andelman et al., 2010; Race et al., 2011), thereby demonstrating that the hippocampus is critical to this process. Similarly an fMRI study found the hippocampus to be active during scene construction in healthy control participants, providing further evidence for the theory (Hassabis et al., 2007b).

However, it is not currently known how scene construction contributes to, and interacts with hippocampal representations of episodic memories. In my final experiment (Chapter 7), I investigated the neural basis of boundary extension in order to better characterise the hippocampal role in scene

construction, so that we might begin to understand the mechanisms by which this process contributes to the representation of episodic memory.

Boundary extension is a cognitive phenomenon whereby participants reliably remember seeing more of a scene than was present in the physical input. This effect is thought to depend on the automatic, implicit construction of scenes beyond the border of a given view, which leads to the subsequent boundary extension error. A recent study demonstrated that amnesic patients with selective bilateral hippocampal lesions showed reduced boundary extension, providing support for the idea that boundary extension depends on hippocampal scene construction processes. However, it was not possible to determine whether the attenuated boundary extension could be attributed to scene construction deficits when the original scene was physically present, or some other effect at the time of the boundary extension error. Here, I used a standard whole brain fMRI paradigm in order to investigate the neural correlates of boundary extension, thereby expanding our knowledge of the role of the hippocampus in automatic scene construction. I particularly focussed on activity during the initial presentation of scenes, in order to test the hypothesis that the hippocampus is involved in the automatic construction of scenes beyond the edges of a given view.

Using a standard univariate fMRI approach, I found that bilateral hippocampus and posterior parahippocampal cortex (PHC) both displayed increased activity on trials where boundary extension was present compared to those where it was not. Importantly this contrast specifically looked at the

activity elicited by the first scene presentation, prior to the behavioural expression of the boundary extension error. Thus, this activity can be specifically attributed to the active extrapolation of a scene beyond the given view. A DCM analysis revealed that the PHC activity was actively modulated by the hippocampus during boundary extension, thereby suggesting that the hippocampus plays the key role in driving this effect. A second analysis revealed that primary visual cortex displayed adaptation effects consistent with changes in subjective perception of the scenes. Specifically, the BOLD adaptation was measured between the first and second scene presentation on each trial, and it was found that trials where the second scene was perceived to be "closer" displayed less adaptation than those where the second scene was perceived to be "the same". This is precisely the pattern of results one would expect if the scene were physically displayed as "closer" or the "same", but notably, the scene pairs were always physically identical in this study. Thus, this effect can only be due to the changes in subjective perception. There must, therefore, be some top-down signal modulating activity within primary visual cortex on boundary extension trials (where the scenes were perceived to be "closer"). A second DCM analysis revealed that this top-down signal was provided by the hippocampus, which drove primary visual cortex activity on boundary extension trials.

Put together these results suggest that the hippocampus plays a crucial role in the automatic anticipation of scene content beyond the physical edges of a scene. I argue that this process is tied to (or driven by) hippocampal scene construction, and this result demonstrates that scene construction processes

can be automatic, rapid, and implicit. The results also show that the extrapolated scene information is channelled back through the scene processing hierarchy as far as primary visual cortex in order to provide predictions about the world beyond the current view. The fact that information about the extended scene boundaries is automatically and rapidly relayed as far back as early visual cortex suggests that scene construction and boundary extension may be a central part of ongoing scene perception in the natural world. Exactly how this process contributes to the formation, storage, and retrieval of individual episodic representations is still not clear, however, and future studies will be required to elucidate this. This issue will be explored in more detail in the next section.

8.6 Conclusions and future directions

This thesis used a combination of fMRI methods in order to investigate the nature of episodic information contained within the human hippocampus. I discovered that it is possible to detect individual episodic memories from patterns of activity within the human hippocampus using MVPA methods. This was also possible with genuine autobiographical memories, regardless of whether those memories were recent or remote, suggesting that remote autobiographical memories always require the hippocampus for rich, vivid episodic recall. Even when the episodes in question contained overlapping elements, it was still possible to detect distinct episodic memory traces from within the hippocampus, demonstrating that episodic memories are coded as unique, conjunctive representations. Within the individual subfields of the hippocampus, I found that regions CA1 and CA3 were particularly

important for the representation of unique, conjunctive episodic representations. In contrast, only CA3 displayed any evidence for pattern completion across overlapping episodes. Furthermore, the degree of distinct episodic information within region CA3, and even the anatomical size of CA3, predicted individual differences in subjective episodic distinctiveness. These latter results demonstrate that differences in episodic representation and even subfield anatomical structure can directly influence the way that we perceive our own memories. Finally, I found that the hippocampus plays a role in the automatic extrapolation of scenes beyond the given view, consistent with a rapid, automatic scene construction process. This result re-emphasises the crucial role of the hippocampus in scene construction, prompting further study into the relationship between scene construction and episodic representation. As a whole, this set of results provides novel empirical data regarding the nature of episodic representation in the human hippocampus. However, there are many questions that cannot be resolved from these data alone. Below I will outline some of the questions raised by the results of this thesis, and suggest future avenues for research that may help to address these questions.

8.6.1 How do memory traces evolve over time?

Using a cross-sectional study of two different time-points (two weeks and 10 years), I have shown that both recent and remote autobiographical memories are represented within the hippocampus. I also found that the posterior hippocampus showed an increase in the strength of representation for the remote memories compared to the recent, thereby mirroring the

results in the cortical regions studied. This demonstrates that there is some form of transformation in the nature of the memory trace over time. However, it does not allow us to conclude exactly what this transformation may be. Does the hippocampal memory trace remain exactly the same in terms of the underlying neuronal population? Or does it change gradually over time, recruiting new neuronal populations, and losing others? This could eventually lead to a memory trace that overlapped with the original trace, but was nevertheless distinct, or could even lead to the memory being represented by a completely different neuronal population. The only way to resolve this important question is to use a longitudinal design, measuring the pattern of activation expressed for the same set of memories at different time-points. This would allow one to explicitly compare the multi-voxel patterns expressed for each memory at the different points in time, and determine whether there is any detectable change in those patterns.

8.6.2 What functional dissociations account for the differences in episodic memory representations along the anterior-posterior axis of the hippocampus?

Besides the functional differentiation between the hippocampal subfields (which I discuss further below), another commonly observed functional dissociation is that of the posterior and anterior portions of the hippocampus (Moser and Moser, 1998; Maguire et al., 2000; Gilboa et al., 2004; Kahn et al., 2008; Fanselow and Dong, 2010; Poppenk and Moscovitch, 2011). This was also the case in this thesis, where I found that the posterior hippocampus displayed an increase in episodic representation for remote

compared to recent memories, whereas the anterior hippocampus displayed no difference between the two conditions.

The posterior and body of the human hippocampus are often associated with spatial memory and spatial representation (e.g. Moser and Moser, 1998; Maguire et al., 2000), while the anterior hippocampus is more often linked to episodic and autobiographical memory (Svoboda et al., 2006). However, the story is more complex than this, as activity relating to spatial processing has been found in anterior hippocampus (Viard et al., 2011), while I have clearly demonstrated that the posterior hippocampus contains distinct representations of episodic memories (Experiments 1 and 2). Thus, both anterior and posterior hippocampus have been associated with both episodic and spatial memory, making it unlikely that this particular functional dissociation can adequately explain the functional dissociation. Nevertheless, the fact remains that the anterior and posterior hippocampus are anatomically distinct in terms of the pattern connectivity with the rest of the brain (Kahn et al., 2008; Poppenk and Moscovitch, 2011), suggesting that they are likely to be involved in distinct (if potentially overlapping) functions. As described above, I found that the posterior hippocampus in particular displayed an increase in episodic representation for remote memories, mirroring the results in neocortex. I believe that this result is likely to be an important part of the puzzle, and that discovering the underlying reason for this functional dissociation may help to resolve this debate.

8.6.3 What are the precise roles of the hippocampal subfields in episodic memory?

I have demonstrated that the episodic representations found within the individual hippocampal subfields are consistent with computational models of hippocampal function. This is an important first step in demonstrating that these theoretical models offer a useful explanatory framework for understanding complex episodic memories. However, there are still many questions to be answered about the role of the subfields. For instance, while I studied episodic representations during the retrieval of well-learned episodes, what would we expect to see during the encoding of episodic memories? Computational models would suggest that the DG and CA3 ought to support the distinct, pattern-separated representations of episodes during encoding, while CA1 and the subiculum may not. A more general question is the nature of the memory traces in region CA1, and how quickly these become established. Theoretical models generally agree that distinct memory traces are rapidly formed within CA3, and that connections between CA3 and CA1 lead to the formation of distinct representations within CA1 also (Rolls, 2010; O'Reilly et al., 2011). However, it is not clear how quickly the CA1 representation is expected to form. Another important question is exactly how high-level goals and task demands influence the computations and representations of the subfields. For instance, what differences might we expect to see between a task emphasising pattern separation, and a task emphasising pattern completion? In such a case, the same set of stimuli may lead to very different representations within the subfields. Future studies using MVPA and high-resolution fMRI may

provide answers to such questions, and allow us to develop more evidence-based computational models of episodic memory.

8.6.4 What roles do the hippocampal subfields play in individual differences in episodic memory?

I found a striking relationship between individual differences in CA3 anatomical size, the amount of episodic information contained within CA3, and the perceived distinctiveness of four overlapping episodes. This relationship was captured by a mediation analysis, whereby increases in CA3 size caused increases in the distinctiveness of CA3 episodic representations, which in turn caused a decrease in how aware the subjects were of the conflicting, overlapping memories. Thus, CA3 anatomy and function were found to have a profound influence on the way that we perceive our own memories. This finding provokes several further questions regarding the relationship between CA3 and individual differences in episodic memory. First, is there a link between this kind of subjective episodic distinctiveness and more objective measures of pattern separation (e.g. Lacy et al., 2011)? If there is, do both processes depend on anatomical and functional differences within region CA3? Second, why is it that increased CA3 size leads to greater distinctiveness in the underlying episodic information? I suggest two plausible explanations for this relationship. First, it might simply be that the increased size is due to a greater number of neurons within CA3, which in turn leads to increased functional capacity. More specifically, more neurons may allow for a greater capacity to store similar representations (overlapping episodes) as distinct

representations. Second, instead of a greater number of neurons, increased CA3 size may reflect an increase in the connectivity between the CA3 neuronal assemblies. Such increased collateral connectivity would in theory allow the CA3 auto-associator architecture to maintain more distinct episodic representations. These are two highly simplistic hypotheses either of which could potentially account for this relationship. Alternatively, the relationship could be due to a mixture of these two hypotheses, or some other, more complex change in neuronal architecture or connectivity. Future studies will be required to determine which explanation is correct. Finally, what factors are involved in these anatomical, functional, and cognitive differences between individuals? For instance, are there genetic factors that determine CA3 anatomy? Or are these differences better explained by environmental factors? Given the demonstration of hippocampal plasticity in taxi drivers (Maguire et al., 2000; Woollett and Maguire, 2011), one particularly interesting question is whether it is possible to selectively increased the size of region CA3. For example, could extensive training on a pattern completion or pattern separation task lead to selective structural differences in CA3?

Overall, the study of the functional role of the human hippocampal subfields is still young, but it is already clear that the different subfields play distinct roles in memory and cognition. It now appears to be the case that the anatomy and function of specific subfields can explain certain mnemonic differences found between individuals. Many questions remain regarding the nature of this relationship, and future studies will no doubt help to elucidate some of the issues raised above.

8.6.5 How do scene construction and boundary extension contribute to the representation of episodic memories at the neural level?

In Experiment 5 I demonstrated that the hippocampus is actively involved in boundary extension, which is dependent on scene construction processes. However, it is currently not clear exactly how processes such as scene construction and boundary extension contribute to the representation of episodic memories. The data in my thesis do not directly speak to this issue, as each study investigated either the representation of episodic memories, or scene construction processes, and never the direct relationship between them. Part of the problem is that it is not clear exactly how we should conceptualise the relationship between a process such as scene construction, and a hippocampal representation such as an episodic memory. I believe that it will be critical to formalise scene construction in terms of a computational model of the hippocampus. By developing such a model, it will be possible to clarify the nature of the representations within the hippocampus, and to define how scene construction contributes to these representations. It would also generate specific, quantitative predictions which could be empirically tested using approaches such as those used throughout this thesis. Finally, by putting scene construction on a more formal footing in this way, it would be possible to integrate it with other computations such as pattern completion and pattern separation in order to build a more complete model of episodic memory. The only extant model that currently tries to link some of these strands together is the model by Byrne et al. (2007), which provides a promising account of the neural machinery underlying both spatial memory and spatial imagery. However, it is not entirely clear how this

model can account for the construction of entirely novel imagined scenes (as opposed to specific, previously experienced environments). Furthermore, the model addresses spatial memory and spatial imagination without specifically linking these processes to episodic memory. Further work specifically aimed at addressing these issues will be required in order to fully understand the neural basis of scene construction and its relation to episodic memory.

8.6.6 Final conclusions

In this thesis I have provided novel evidence regarding the nature of episodic representations in the human hippocampus, as well as their instantiation in hippocampal subfields. While these results provide an important first step in the task of characterising hippocampal representations, many questions remain, some of which I have discussed above. Hopefully, my contribution to the development and application of MVPA to the hippocampus and its subfields will allow future studies to address these questions. In so doing, I have no doubt that we will be able to develop a more complete understanding of how the brain manages the complex storage and retrieval problems posed by episodic memory, and how these neural signals allow us to re-experience the past in such rich and vivid detail.

9 References

Addis DR, Moscovitch M, Crawley AP, McAndrews MP (2004) Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. Hippocampus 14:752–762.

Addis DR, Wong AT, Schacter DL (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45:1363–1377.

Allwein E, Shapire R, Singer Y (2000) Reducing multiclass to binary: a unifying approach for margin classifiers. Journal of Machine Learning Research 1:113–141.

Amaral DG (1999) Introduction: what is where in the medial temporal lobe? Hippocampus 9:1–6.

Andelman F, Hoofien D, Goldberg I, Aizenstein O, Neufeld MY (2010) Bilateral hippocampal lesion and a selective impairment of the ability for mental time travel. Neurocase 16:426–435.

Andersen P, Morris R, Amaral D, Bliss T, O'Keefe J (2006) The Hippocampus Book, 1st ed. New York, USA: Oxford University Press.

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. Neuroimage 13:903–919.

Ashburner J (2007) A fast diffeomorphic image registration algorithm. NeuroImage 38:95–113.

Bakker A, Kirwan CB, Miller M, Stark CEL (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. Science 319:1640–1642.

Bartlett FC (1932) Remembering. Cambridge, UK: Cambridge University Press.

Bartsch T, Döhring J, Rohr A, Jansen O, Deuschl G (2011) CA1 neurons in the human hippocampus are critical for autobiographical memory, mental time travel, and autonoetic consciousness. Proc Natl Acad Sci USA 108:17562–17567.

Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV (2012) Decoding the neural representation of affective states. Neuroimage 59:718–727.

Bayley PJ, Gold JJ, Hopkins RO, Squire LR (2005) The neuroanatomy of remote memory. Neuron 46:799–810.

Bird CM, Burgess N (2008) The hippocampus and memory: insights from spatial processing. Nat Rev Neurosci 9:182–194.

Bliss TV, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361:31–39.

Bonnici HM, Kumaran D, Chadwick MJ, Weiskopf N, Hassabis D, Maguire EA (2012) Decoding representations of scenes in the medial temporal lobes. Hippocampus 22:1143–1153.

Bontempi B, Laurent-Demir C, Destrade C, Jaffard R (1999) Time-dependent reorganization of brain circuitry underlying long-term memory storage. Nature 400:671–675.

Botzung A, Denkova E, Manning L (2008) Experiencing past and future personal events: functional neuroimaging evidence on the neural bases of mental time travel. Brain Cogn 66:202–212.

Burgess N, Barry C, O'Keefe J (2007) An oscillatory interference model of grid cell firing. Hippocampus 17:801–812.

Burgess N, Maguire EA, O'Keefe J (2002) The human hippocampus and spatial and episodic memory. Neuron 35:625–641.

Byrne P, Becker S, Burgess N (2007) Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol Rev 114(2):340-375.

Cabeza R, Ciaramelli E, Olson IR, Moscovitch M (2008) The parietal cortex and episodic memory: an attentional account. Nat Rev Neurosci 9:613–625.

Cajal SR (1911) Histologie du Systeme Nerveux de l'Homme et des Vertebres. Paris: Maloine.

Candel I, Merckelbach H, Houben K, Vandyck I (2004) How children remember neutral and emotional pictures: boundary extension in children's scene memories. Am J Psychol 117:249–257.

Carr VA, Rissman J, Wagner AD (2010) Imaging the human medial temporal lobe with high-resolution fMRI. Neuron 65:298–308.

Chang C, Lin C (2011) LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology 2:Article 17.

Cipolotti L, Bird CM (2006) Amnesia and the hippocampus. Curr Opin Neurol 19:593–598.

Clark RE, Broadbent NJ, Squire LR (2005a) Hippocampus and remote spatial memory in rats. Hippocampus 15:260–272.

Clark RE, Broadbent NJ, Squire LR (2005b) Impaired remote spatial memory after hippocampal lesions despite extensive training beginning early in life. Hippocampus 15:340–346.

Clayton NS, Bussey TJ, Dickinson A (2003) Can animals recall the past and plan for the future? Nat Rev Neurosci 4:685–691.

Cohen NJ, Eichenbaum H (1993) Memory, amnesia and the hippocampal system. Cambridge, USA: MIT press.

Conway MA, Pleydell-Pearce CW (2000) The construction of autobiographical memories in the self-memory system. Psychol Rev 107:261–288.

Corkin S (2002) What's new with the amnesic patient H.M.? Nat Rev Neurosci 3:153–160.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261–270.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

deCharms RC, Zador A (2000) Neural representation and the cortical code. Annu Rev Neurosci 23:613–647.

Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19:430–441.

Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. Neuroimage 21:757–767.

Derdikman D, Moser EI (2010) A manifold of spatial maps in the brain. Trends Cogn Sci 14:561–569.

Diana RA, Yonelinas AP, Ranganath C (2008) High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. Hippocampus 18:536–541.

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302.

Diedrichsen J, Ridgway GR, Friston KJ, Wiestler T (2011) Comparing the similarity and spatial structure of neural representations: a pattern-component model. Neuroimage 55:1665–1678.

Dietterich TD, Bakiri G (1994) Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research 2:263–286.

Doeller CF, Barry C, Burgess N (2010) Evidence for grid cells in a human memory network. Nature 463:657–661.

Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb Cortex 19:2269–2280.

Duda OR, Hart PE, Stork DG (2001) Pattern Classification. New York, USA: Wiley.

Dudai Y (2004) The neurobiology of consolidations, or, how stable is the engram? Annu Rev Psychol 55:51–86.

Duncan K, Ketz N, Inati SJ, Davachi L (2012) Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus 22:389–398.

Duong TQ, Kim DS, Uğurbil K, Kim SG (2001) Localized cerebral blood flow response at submillimeter columnar resolution. Proc Natl Acad Sci USA 98:10904–10909.

Duvernoy HM (1999) The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply, 2nd ed. New York, USA: Springer.

Duvernoy HM (2005) The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections with MRI, 3rd ed. New York, USA: Springer.

Eacott MJ, Easton A (2010) Episodic memory in animals: remembering which occasion. Neuropsychologia 28(8):2273-2280.

Eichenbaum H (2000) Hippocampus: mapping or memory? Curr Biol 10:R785–R787.

Eichenbaum H (2004) Hippocampus: cognitive processes and neural representations that underlie declarative memory. Neuron 44:109–120.

Eichenbaum H, Stewart C, Morris RG (1990) Hippocampal representation in place learning. J Neurosci 10:3531–3542.

Ekstrom AD, Bazih AJ, Suthana NA, Al-Hakim R, Ogura K, Zeineh M, Burggren AC, Bookheimer SY (2009) Advances in high-resolution imaging and computational unfolding of the human hippocampus. Neuroimage 47:42–49.

Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. Nature 425:184–188.

Eldridge LL, Engel SA, Zeineh MM, Bookheimer SY, Knowlton BJ (2005) A dissociation of encoding and retrieval processes in the human hippocampus. J Neurosci 25:3280–3286.

Epstein R (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci 12:388–396.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601.

Epstein RA, Morgan LK (2012) Neural responses to visual scenes reveals inconsistencies between fMRI adaptation and multivoxel pattern analysis. Neuropsychologia 50:530–543.

Etzel JA, Gazzola V, Keysers C (2009) An introduction to anatomical ROI-based fMRI classification analysis. Brain Res 1282:114–125.

Fanselow MS, Dong H-W (2010) Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65:7–19.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004) Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11–22.

Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner JT, Penny WD (2004) Human brain function. New York, USA: Elsevier Academic Press.

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not columns. J Neurosci 31:4792–4804.

Frey U, Morris RG (1998) Synaptic tagging: implications for late maintenance of hippocampal long-term potentiation. Trends Neurosci 21:181–188.

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138.

Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J (2008) Bayesian decoding of brain images. Neuroimage 39:181–205.

Friston KJ, Frith CD, Frackowiak RS, Turner R (1995) Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage 2:166–172.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302.

Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. Science 322:96–101.

Gilboa A, Winocur G, Grady CL, Hevenor SJ, Moscovitch M (2004) Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. Cereb Cortex 14:1214–1225.

Goense JBM, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol 18:631–640.

Goshen I, Brodsky M, Prakash R, Wallace J, Gradinaru V, Ramakrishnan C, Deisseroth K (2011) Dynamics of retrieval strategies for remote memories. Cell 147:678–689.

Gottesman CV, Intraub H (2002) Surface construal and the mental representation of scenes. Journal of Experimental Psychology: Human Perception and Performance 28:589–599.

Gregory RL (1968) Perceptual Illusions and Brain Models. Philos Trans R Soc Lond B Biol Sci 171:279–296.

Gregory RL (1980) Perceptions as Hypotheses. Philos Trans R Soc Lond B Biol Sci 290:181–197.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23.

Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. Vision Res 41:1409–1422.

Guyon I, Elisseeff A (2003) An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3:1157–1182.

Hackert VH, den Heijer T, Oudkerk M, Koudstaal PJ, Hofman A, Breteler MMB (2002) Hippocampal head size associated with verbal memory performance in nondemented elderly. Neuroimage 17:1365–1372.

Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. Nature 436:801–806.

Hamming RW (1950) Error-detecting and error-correcting. Bell System Technical Journal 29:147–160.

Hannula DE, Tranel D, Cohen NJ (2006) The long and the short of it: relational memory impairments in amnesia, even at short lags. J Neurosci 26:8352–8359.

Hanseeuw BJ, Van Leemput K, Kavec M, Grandin C, Seron X, Ivanoiu A (2011) Mild cognitive impairment: differential atrophy in the hippocampal subfields. AJNR Am J Neuroradiol 32:1658–1661.

Hartley T, Bird CM, Chan D, Cipolotti L, Husain M, Vargha-Khadem F, Burgess N (2007) The hippocampus is required for short-term topographical memory in humans. Hippocampus 17:34–48.

Hartley T, Maguire EA, Spiers HJ, Burgess N (2003) The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. Neuron 37:877–888.

Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA (2009) Decoding neuronal ensembles in the human hippocampus. Curr Biol 19:546–554.

Hassabis D, Kumaran D, Maguire EA (2007b) Using imagination to understand the neural basis of episodic memory. J Neurosci 27:14365–14374.

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007a) Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci USA 104:1726–1731.

Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. Trends Cogn Sci 11:299–306.

Hassabis D, Maguire EA (2009) The construction system of the brain. Philos Trans R Soc Lond, B, Biol Sci 364:1263–1271.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Haynes J-D, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Heeger DJ, Ress D (2002) What does fMRI tell us about neuronal activity? Nat Rev Neurosci 3:142–151.

Henderson JM, Hollingworth A (1999) High-level scene perception. Annu Rev Psychol 50:243–271.

Hodges JR, Patterson K, Oxbury S, Funnell E (1992) Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. Brain 115 (Pt 6):1783–1806.

Hoscheidt SM, Nadel L, Payne J, Ryan L (2010) Hippocampal activation during retrieval of spatial context from episodic and semantic memory. Behav Brain Res 212:121–132.

Hsu CW, Lin CJ (2002) A comparison of methods for multi-class support vector machines. IEEE Transaction on Neural Networks 13:415–425.

Hubel DH (1963) The Visual Cortex Of The Brain. Sci Am 209:54–62.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image Distortion Correction in fMRI: A Quantitative Evaluation. Neuroimage 16:217–240.

Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. AJNR Am J Neuroradiol 19:659–671.

Intraub H (2004) Anticipatory spatial representation of 3D regions explored by sighted observers and a deaf-and-blind-observer. Cognition 94:19–37.

Intraub H (2012) Rethinking visual scene perception. Wiley Interdisciplinary Reviews: Cognitive Science 3:117–127.

Intraub H, Dickinson C (2008) False memory 1/20th of a second later: what the early onset of boundary extension reveals about perception. Psychol Sci 19:1007–1014.

Intraub H, Gottesman CV, Willey EV, Zuk IJ (1996) Boundary Extension for Briefly Glimpsed Photographs: Do Common Perceptual Processes Result in Unexpected Memory Distortions? Journal of Memory and Language 35:118–134.

Intraub H, Gottesman CV, Bills AJ (1998) Effects of perceiving and imagining scenes on memory for pictures. J Exp Psychol Learn Mem Cogn 24:186–201.

Intraub H, Richardson M (1989) Wide-angle memories of close-up scenes. J Exp Psychol Learn Mem Cogn 15:179–187.

Jezzard P, Matthews PM, Smith SM eds. (2003) Functional MRI: An Introduction to Methods. New York, USA: Oxford University Press.

Kahn I, Andrews-Hanna JR, Vincent JL, Snyder AZ, Buckner RL (2008) Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. J Neurophysiol 100:129–139.

Kahnt T, Heinzle J, Park SQ, Haynes J-D (2011) Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. Neuroimage 56:709–715.

Kamitani Y, Sawahata Y (2010) Spatial smoothing hurts localization but not information: pitfalls for brain mappers. Neuroimage 49:1949–1952.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355.

Kim JJ, Fanselow MS (1992) Modality-specific retrograde amnesia of fear. Science 256:675–677.

Kirwan CB, Bayley PJ, Galván VV, Squire LR (2008) Detailed recollection of remote autobiographical memory after damage to the medial temporal lobe. Proc Natl Acad Sci USA 105:2676–2680.

Kirwan CB, Jones CK, Miller MI, Stark CEL (2007) High-resolution fMRI investigation of the medial temporal lobe. Hum Brain Mapp 28:959–966.

Klein SB, Loftus J, Kihlstrom JJ (2002) Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. Soc Cogn 20:353–379.

Kopelman MD, Wilson BA, Baddeley AD (1989) The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. J Clin Exp Neuropsychol 11:724–744.

Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. Science 293:1506–1509.

Kriegeskorte N (2011) Pattern-information analysis: from stimulus decoding to computational-model testing. Neuroimage 56:411–421.

Kriegeskorte N, Cusack R, Bandettini P (2010) How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? Neuroimage 49:1965–1976.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci USA 103:3863–3868.

Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540.

Ku S-pi, Gretton A, Macke J, Logothetis NK (2008) Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. Magn Reson Imaging 26:1007–1014.

Kumaran D, Maguire EA (2006) An unexpected sequence of events: mismatch detection in the human hippocampus. PLoS Biol 4:e424.

Kumaran D, Maguire EA (2009) Novelty signals: a window into hippocampal information processing. Trends Cogn Sci 13:47–54.

LaConte S, Strother S, Cherkassky V, Anderson J, Hu X (2005) Support vector machines for temporal classification of block design fMRI data. Neuroimage 26:317–329.

Lacy JW, Yassa MA, Stark SM, Muftuler LT, Stark CEL (2011) Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. Learn Mem 18:15–18.

Lashley K (1950) In search of the engram. Symp Soc Exp Biol 4:454–482.

Lee ACH, Buckley MJ, Pegman SJ, Spiers H, Scahill VL, Gaffan D, Bussey TJ, Davies RR, Kapur N, Hodges JR, Graham KS (2005a) Specialization in the medial temporal lobe for processing of objects and scenes. Hippocampus 15:782–797.

Lee ACH, Bussey TJ, Murray EA, Saksida LM, Epstein RA, Kapur N, Hodges JR, Graham KS (2005b) Perceptual deficits in amnesia: challenging the medial temporal lobe “mnemonic” view. Neuropsychologia 43:1–11.

Lee ACH, Scahill VL, Graham KS (2008) Activating the medial temporal lobe during oddity judgment for faces and scenes. Cereb Cortex 18:683–696.

Lee I, Yoganarasimha D, Rao G, Knierim JJ (2004) Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. Nature 430:456–459.

Leutgeb JK, Leutgeb S, Moser M-B, Moser EI (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315:961–966.

Leutgeb S, Leutgeb JK, Treves A, Moser M-B, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. Science 305:1295–1298.

Lever C, Wills T, Cacucci F, Burgess N, O’Keefe J (2002) Long-term plasticity in hippocampal place-cell representation of environmental geometry. Nature 416:90–94.

Levine B, Svoboda E, Hay JF, Winocur G, Moscovitch M (2002) Aging and autobiographical memory: dissociating episodic from semantic retrieval. Psychol Aging 17:677–689.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869–878.

Lorente De No R (1933) Studies on the structure of the cerebral cortex. I. The Area Entorhinalis. J Psychol Neurol 45:381–438.

Lorente De No R (1934) Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. J Psychol Neurol 46:113–177.

Magri C, Schridde U, Murayama Y, Panzeri S, Logothetis NK (2012) The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. J Neurosci 32:1395–1407.

Maguire EA (2001) Neuroimaging studies of autobiographical event memory. Philos Trans R Soc Lond, B, Biol Sci 356:1441–1451.

Maguire EA, Frith CD (2003) Lateral asymmetry in the hippocampal response to the remoteness of autobiographical memories. J Neurosci 23:5302–5307.

Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci USA 97:4398–4403.

Maguire EA, Henson RN, Mummery CJ, Frith CD (2001) Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. Neuroreport 12:441–444.

Maguire EA, Nannery R, Spiers HJ (2006) Navigation around London by a taxi driver with bilateral hippocampal lesions. Brain 129:2894–2907.

Malykhin NV, Lebel RM, Coupland NJ, Wilman AH, Carter R (2010) In vivo quantification of hippocampal subfields using 4.7 T fast spin echo imaging. Neuroimage 49:1224–1230.

Mansfield P (1977) Multi-planar imaging using NMR spin echoes. Journal of Physics C 10:L55–L58.

Marr D (1971) Simple memory: a theory for archicortex. Philos Trans R Soc Lond, B, Biol Sci 262:23–81.

Marshall GA, Kaufer DI, Lopez OL, Rao GR, Hamilton RL, DeKosky ST (2004) Right prosubiculum plaque density correlates with anosognosia in Alzheimer’s disease. J Neurol Neurosurg Psychiatry 75:1396-1400.

Marslen-Wilson WD, Teuber HL (1975) Memory for remote events in anterograde amnesia: recognition of public figures from newsphotographs. Neuropsychologia 13:353–364.

Martin VC, Schacter DL, Corballis MC, Addis DR (2011) A role for the hippocampus in encoding simulations of future events. Proc Natl Acad Sci USA 108:13858–13863.

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419–457.

McKenzie S, Eichenbaum H (2011) Consolidation and reconsolidation: two lives of memories? Neuron 71:224–233.

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser M-B (2006) Path integration and the neural basis of the "cognitive map." Nat Rev Neurosci 7:663–678.

Menon RS, Ogawa S, Hu X, Strupp JP, Anderson P, Uğurbil K (1995) BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. Magn Reson Med 33:453–459.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 53:103–118.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X (2004) Learning to Decode Cognitive States from Brain Images. Mach Learn 57:145–175.

Montaldi D, Mayes AR (2011) Familiarity, recollection and medial temporal lobe function: an unresolved issue. Trends Cogn Sci 15:339–340.

Morcom AM, Friston KJ (2012) Decoding episodic memory in ageing: a Bayesian analysis of activity patterns predicting memory. Neuroimage 59:1772–1782.

Moreno H, Wu WE, Lee T, Brickman A, Mayeux R, Brown TR, Small SA (2007) Imaging the Abeta-related neurotoxicity of Alzheimer disease. Arch Neurol 64:1467–1477.

Morris RG, Garrud P, Rawlins JN, O'Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. Nature 297:681–683.

Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, McAndrews MP, Levine B, Black S, Winocur G, Nadel L (2005) Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. J Anat 207:35–66.

Moser EI, Kropff E, Moser M-B (2008) Place cells, grid cells, and the brain's spatial representation system. Annu Rev Neurosci 31:69–89.

Moser MB, Moser EI (1998) Functional differentiation in the hippocampus. Hippocampus 8:608–619.

Mueller SG, Schuff N, Yaffe K, Madison C, Miller B, Weiner MW (2010) Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. Hum Brain Mapp 31:1339–1347.

Mueller SG, Stables L, Du AT, Schuff N, Truran D, Cashdollar N, Weiner MW (2007) Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4T. Neurobiol Aging 28:719–726.

Mugler JP 3rd, Bao S, Mulkern RV, Guttmann CR, Robertson RL, Jolesz FA, Brookeman JR (2000) Optimized single-slab three-dimensional spin-echo MR imaging of the brain. Radiology 216:891–899.

Mullally SL, Intraub H, Maguire EA (2012) Attenuated Boundary Extension Produces a Paradoxical Memory Advantage in Amnesic Patients. Curr Biol 22:261-268.

Mullally SL, Maguire EA (2011) A new role for the parahippocampal cortex in representing space. J Neurosci 31:7441–7449.

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643.

Murray SO, Boyaci H, Kersten D (2006) The representation of perceived angular size in human primary visual cortex. Nat Neurosci 9:429–434.

Nadel L, Land C (2000) Memory traces revisited. Nat Rev Neurosci 1:209–212.

Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. Curr Opin Neurobiol 7:217–227.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902–915.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25.

Nieuwenhuis ILC, Takashima A (2011) The role of the ventromedial prefrontal cortex in memory consolidation. Behav Brain Res 218:325–334.

Niki K, Luo J (2002) An fMRI study on the time-limited role of the medial temporal lobe in long-term topographical autobiographic memory. J Cogn Neurosci 14:500–507.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 34:171–175.

O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. Oxford University Press, USA.

O'Reilly RC, Bhattacharyya R, Howard MD, Ketz N (2011) Complementary Learning Systems. Cogn Sci (in press).

Ogawa S, Lee TM (1990) Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. Magn Reson Med 16:9–18.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci USA 87:9868–9872.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci USA 89:5951–5955.

Okuda J, Fujii T, Ohtake H, Tsukiura T, Tanji K, Suzuki K, Kawashima R, Fukuda H, Itoh M, Yamadori A (2003) Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. Neuroimage 19:1369–1380.

Ollinger JM, Corbetta M, Shulman GL (2001) Separating Processes within a Trial in Event-Related Functional MRI: II. Analysis. Neuroimage 13:218–229.

Olson IR, Page K, Moore KS, Chatterjee A, Verfaellie M (2006) Working memory for conjunctions relies on the medial temporal lobe. J Neurosci 26:4596–4601.

Ono T, Nakamura K, Fukuda M, Tamura R (1991) Place recognition responses of neurons in monkey hippocampus. Neurosci Lett 121:194–198.

Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011) A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage 56:593–600.

Op de Beeck HP (2010a) Probing the mysterious underpinnings of multi-voxel fMRI analyses. Neuroimage 50:567–571.

Op de Beeck HP (2010b) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage 49:1943–1948.

Park S, Intraub H, Yi D-J, Widders D, Chun MM (2007) Beyond the edges of a view: boundary extension in human scene-selective visual cortex. Neuron 54:335–342.

Peelen MV, Atkinson AP, Vuilleumier P (2010) Supramodal representations of perceived emotions in the human brain. J Neurosci 30:10127–10134.

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199–S209.

Piefke M, Weiss PH, Zilles K, Markowitsch HJ, Fink GR (2003) Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. Brain 126:650–668.

Piolino P, Giffard-Quillon G, Desgranges B, Chételat G, Baron J-C, Eustache F (2004) Re-experiencing old memories via hippocampus: a PET study of autobiographical memory. Neuroimage 22:1371–1383.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Poppenk J, Moscovitch M (2011) A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. Neuron 72:931–937.

Preacher KJ, Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 36:717–731.

Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 40:879–891.

Preston AR, Bornstein AM, Hutchinson JB, Gaare ME, Glover GH, Wagner AD (2010) High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. J Cogn Neurosci 22:156–173.

Quinn PC, Intraub H (2007) Perceiving “outside the box” occurs early in development: evidence for boundary extension in three- to seven-month-old infants. Child Dev 78:324–334.

Race E, Keane MM, Verfaellie M (2011) Medial Temporal Lobe Damage Causes Deficits in Episodic Memory and Episodic Future Thinking Not Attributable to Deficits in Narrative Construction. J Neurosci 31:10262–10269.

Redondo RL, Morris RGM (2011) Making memories last: the synaptic tagging and capture hypothesis. Nat Rev Neurosci 12:17–30.

Rekkas PV, Constable RT (2005) Evidence that autobiographic memory retrieval does not become independent of the hippocampus: an fMRI study contrasting very recent with remote events. J Cogn Neurosci 17:1950–1961.

Rempel-Clower NL, Zola SM, Squire LR, Amaral DG (1996) Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci 16:5233–5255.

Rissman J, Greely HT, Wagner AD (2010) Detecting individual memories through the neural decoding of memory states and past experience. Proc Natl Acad Sci USA 107:9849–9854.

Rolls ET (2010) A computational theory of episodic memory formation in the hippocampus. Behav Brain Res 215:180–196.

Rosenbaum RS, Gilboa A, Levine B, Winocur G, Moscovitch M (2009) Amnesia as an impairment of detail generation and binding: evidence from personal, fictional, and semantic narratives in K.C. Neuropsychologia 47:2181–2187.

Rosenbaum RS, Moscovitch M, Foster JK, Schnyer DM, Gao F, Kovacevic N, Verfaellie M, Black SE, Levine B (2008) Patterns of autobiographical memory loss in medial-temporal lobe amnesic patients. J Cogn Neurosci 20:1490–1506.

Rugg MD, Otten LJ, Henson RNA (2002) The neural basis of episodic memory: evidence from functional neuroimaging. Philos Trans R Soc Lond, B, Biol Sci 357:1097–1110.

Ryan L, Nadel L, Keil K, Putnam K, Schnyer D, Trouard T, Moscovitch M (2001) Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people. Hippocampus 11:707–714.

Salwiczek LH, Watanabe A, Clayton NS (2010) Ten years of research into avian models of episodic-like memory and its implications for developmental and comparative cognition. Behav Brain Res 215:221–234.

Schacter DL, Norman KA, Koutstaal W (1998) The cognitive neuroscience of constructive memory. Annu Rev Psychol 49:289–318.

Schnider A (2003) Spontaneous confabulation and the adaptation of thought to ongoing reality. Nat Rev Neurosci 4:662–671.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatr 20:11–21.

Seamon JG, Schlegel SE, Hiester PM, Landau SM, Blumenthal BF (2002) Misremembering pictured objects: people of all ages demonstrate the boundary extension illusion. Am J Psychol 115:151–167.

Semon R (1923) Mnemonic Psychology. London: Allen & Unwin.

Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr Opin Neurobiol 4:569–579.

Shapley R, Hawken M, Xing D (2007) The dynamics of visual responses in the primary visual cortex. Prog Brain Res 165:21–32.

Sharon T, Moscovitch M, Gilboa A (2011) Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. Proc Natl Acad Sci USA 108:1146–1151.

Shimamura AP, Wickens TD (2009) Superadditive memory strength for item and source recognition: the role of hierarchical relational binding in the medial temporal lobe. Psychol Rev 116:1–19.

Simons JS, Spiers HJ (2003) Prefrontal and medial temporal lobe interactions in long-term memory. Nat Rev Neurosci 4:637–648.

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18:555–586.

Sperandio I, Chouinard PA, Goodale MA (2012) Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. Nat Neurosci 15:540–542.

Spiers HJ, Maguire EA (2006) Thoughts, behaviour, and brain dynamics during navigation in the real world. Neuroimage 31:1826–1840.

Spiers HJ, Maguire EA, Burgess N (2001) Hippocampal amnesia. Neurocase 7:357–382.

Spreng RN, Mar RA, Kim ASN (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. J Cogn Neurosci 21:489–510.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195–231.

Squire LR, Alvarez P (1995) Retrograde amnesia and memory consolidation: a neurobiological perspective. Curr Opin Neurobiol 5:169–177.

Squire LR, Bayley PJ (2007) The neuroscience of remote memory. Curr Opin Neurobiol 17:185–196.

Squire LR, Stark CEL, Clark RE (2004) The medial temporal lobe. Annu Rev Neurosci 27:279–306.

Squire LR, Zola SM (1998) Episodic memory, semantic memory, and amnesia. Hippocampus 8:205–211.

Steinvorth S, Corkin S, Halgren E (2006) Ecphory of autobiographical memories: an fMRI study of recent and remote memory retrieval. Neuroimage 30:285–298.

Steinvorth S, Levine B, Corkin S (2005) Medial temporal lobe structures are needed to re-experience remote autobiographical memories: evidence from H.M. and W.R. Neuropsychologia 43:479–496.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017.

Stephan KE, Penny WD, Moran RJ, den Ouden HEM, Daunizeau J, Friston KJ (2010) Ten simple rules for dynamic causal modeling. Neuroimage 49:3099–3109.

Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. Neuroimage 38:387–401.

Suddendorf T, Busby J (2003) Mental time travel in animals? Trends Cogn Sci 7:391–396.

Suthana N, Ekstrom A, Moshirvaziri S, Knowlton B, Bookheimer S (2011) Dissociations within human hippocampal subregions during encoding and retrieval of spatial information. Hippocampus 21:694–701.

Suthana NA, Ekstrom AD, Moshirvaziri S, Knowlton B, Bookheimer SY (2009) Human hippocampal CA1 involvement during allocentric encoding of spatial information. J Neurosci 29:10512–10519.

Sutherland RJ, Weisend MP, Mumby D, Astur RS, Hanlon FM, Koerner A, Thomas MJ, Wu Y, Moses SN, Cole C, Hamilton DA, Hoesing JM (2001) Retrograde amnesia after hippocampal damage: recent vs. remote memories in two tasks. Hippocampus 11:27–42.

Svoboda E, McKinnon MC, Levine B (2006) The functional neuroanatomy of autobiographical memory: a meta-analysis. Neuropsychologia 44:2189–2208.

Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon C-H, Kim S-G, Tong F (2010) Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. J Neurosci 30:325–330.

Szpunar KK, Watson JM, McDermott KB (2007) Neural substrates of envisioning the future. Proc Natl Acad Sci USA 104:642–647.

Takehara K, Kawahara S, Kirino Y (2003) Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. J Neurosci 23:9897–9905.

Taube JS (2007) The head direction signal: origins and sensory-motor integration. Annu Rev Neurosci 30:181–207.

Teicher MH, Anderson CM, Polcari A (2012) Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. Proc Natl Acad Sci USA 109:E563–E572.

Teyler TJ, DiScenna P (1985) The role of hippocampus in memory: a hypothesis. Neurosci Biobehav Rev 9:377–389.

Tong F (2003) Primary visual cortex and visual awareness. Nat Rev Neurosci 4:219–229.

Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374–391.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM (2007) Schemas and memory consolidation. Science 316:76–82.

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RGM (2011) Schema-dependent gene activation and memory encoding in neocortex. Science 333:891–895.

Tulving E (1972) Episodic and Semantic Memory. In: Organization of Memory (Tulving E, Donaldson W, eds), pp 381–403. New York, USA: Academic.

Tulving E (1983) Elements of Episodic Memory. Oxford, UK: Clarendon.

Tulving E (2002) Episodic memory: from mind to brain. Annu Rev Psychol 53:1–25.

van Kesteren MTR, Fernández G, Norris DG, Hermans EJ (2010) Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. Proc Natl Acad Sci USA 107:7550–7555.

Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009) Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus 19:549–557.

van Strien NM, Widerøe M, van de Berg WDJ, Uylings HBM (2012) Imaging hippocampal subregions with in vivo MRI: advances and limitations. Nat Rev Neurosci 13:70.

Vann SD, Aggleton JP, Maguire EA (2009) What does the retrosplenial cortex do? Nat Rev Neurosci 10:792–802.

Vazdarjanova A, Guzowski JF (2004) Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. J Neurosci 24:6489–6496.

Viard A, Doeller CF, Hartley T, Bird CM, Burgess N (2011) Anterior hippocampus and goal-directed spatial decision making. J Neurosci 31:4613–4621.

Viard A, Piolino P, Desgranges B, Chételat G, Lebreton K, Landeau B, Young A, De La Sayette V, Eustache F (2007) Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: an fMRI study. Cereb Cortex 17:2453–2467.

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. Neuroimage 33:493–504.

Wheeler MA, Stuss DT, Tulving E (1997) Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. Psychol Bull 121:331–354.

Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci USA 97:11125–11129.

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308:873–876.

Winocur G (1990) Anterograde and retrograde amnesia in rats with dorsal hippocampal or dorsomedial thalamic lesions. Behav Brain Res 38:145–154.

Winocur G, Moscovitch M (2011) Memory transformation and systems consolidation. J Int Neuropsychol Soc 17:766–780.

Winocur G, Moscovitch M, Bontempi B (2010) Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions. Neuropsychologia 48:2339–2356.

Winocur G, Moscovitch M, Fogel S, Rosenbaum RS, Sekeres M (2005) Preserved spatial memory after hippocampal lesions: effects of extensive experience in a complex environment. Nat Neurosci 8:273–275.

Wixted JT, Squire LR (2011) The familiarity/recollection distinction does not illuminate medial temporal lobe function: response to Montaldi and Mayes. Trends Cogn Sci 15:340–341.

Wolpert DM, Ghahramani Z (2000) Computational principles of movement neuroscience. Nat Neurosci 3 Suppl:1212–1217.

Woollett K, Maguire EA (2011) Acquiring “the Knowledge” of London’s layout drives structural brain changes. Curr Biol 21:2109–2114.

Yartsev MM, Witter MP, Ulanovsky N (2011) Grid cells without theta oscillations in the entorhinal cortex of bats. Nature 479:103–107.

Yassa MA, Mattfeld AT, Stark SM, Stark CEL (2011) Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. Proc Natl Acad Sci USA 108:8873–8878.

Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31:1116–1128.

Yushkevich PA, Wang H, Pluta J, Das SR, Craige C, Avants BB, Weiner MW, Mueller S (2010) Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. Neuroimage 53:1208–1224.

Zeineh MM, Engel SA, Bookheimer SY (2000) Application of cortical unfolding techniques to functional MRI of the human hippocampal region. Neuroimage 11:668–683.

Zeineh MM, Engel SA, Thompson PM, Bookheimer SY (2003) Dynamics of the hippocampus during encoding and retrieval of face-name pairs. Science 299:577–580.

Zola-Morgan SM, Squire LR (1990) The primate hippocampal formation: evidence for a time-limited role in memory storage. Science 250:288–290.